

DIPLOMARBEIT

Homogenisation of structural breaks in the global ESA CCI Soil Moisture (multi-satellite) climate data record

zur Erlangung des akademischen Grades

Diplom-Ingenieur

im Rahmen des Studiums

Geodäsie und Geoinformation

eingereicht von

Wolfgang Preimesberger

Matrikelnummer 01126962

ausgeführt am Department für Geodäsie und Geoinformation
der Fakultät für Mathematik und Geoinformation der Technischen Universität Wien
Betreuer: Univ.Prof. Dr.rer.nat. Wouter Arnoud Dorigo MSc

Wien, 20.11.2019

(Unterschrift Verfasser/in)

(Unterschrift Betreuer/in)

Abstract

The ESA CCI Soil Moisture COMBINED product is a more than 40 year long data record on global soil moisture for climate studies and applications. It merges soil moisture observations derived from multiple active and passive satellite remote sensing instruments in the microwave domain. Differences in sensor characteristics (such as frequency or polarisation) can cause structural breaks in the product, which are not completely removed during the merging process. These artificially caused discontinuities can adversely affect studies using the long-term data set as they make changes in the observations unrepresentative of long-term changes in reality. Here we compare three adjustment methods in terms of reducing the number of detected breaks in the soil moisture record. We investigate their impact on the data with multiple validation metrics. Their potential (negative) influence is examined by comparing trends in the data before and after homogenisation. We find that all three presented methods can reduce the number of detected breaks in ESA CCI SM. Differences between the methods mainly concern their ability to handle inhomogeneities in variance. Evaluation of the corrected data shows limited impact of homogenisation in terms of quantitative validation metrics. Changes in soil moisture trends due to removing breaks are found in some areas. We find that break correction overall improves the already rather homogeneous data set while preserving its climate describing characteristics. Quantile Category Matching is identified as the preferred method in terms of correcting breaks in ESA CCI SM.

Kurzfassung

ESA CCI Soil Moisture ist ein mehr als 40 Jahre umfassender Datensatz, der den global Bodenwassergehalt beschreibt. Es werden darin Messungen unterschiedlicher aktiver und passiver, satellitengestützter Mikrowellen-Fernerkundungssystemen vereint. Unterschiede in den Sensoreigenschaften (wie Wellenlänge oder Polarisation der gemessenen Strahlung) der kombinierten Messsysteme können zu Inhomogenitäten (“Sprüngen”) im finalen Datensatz führen, die weiterführende Studien negativ beeinflussen können. Diese Arbeit befasst sich mit der Untersuchung von drei Methodiken zur Reduzierung von Inhomogenitäten im ESA CCI SM Datensatz. Ihr Potential zur Verbesserung der Homogenität innerhalb des Datensatzes wird ebenso wie ihre Auswirkungen auf den Datensatz im Bezug auf Validierungsmetriken untersucht. Der Einfluss von Inhomogenitäten und deren Korrektur auf Langzeitstudien wird anhand von Trends im Bodenfeuchtesignal untersucht. Wir kommen zum Schluss, dass alle drei Methoden die Anzahl der detektierbaren Inhomogenitäten im Datensatz reduzieren, wobei Unterschiede im Potential zur Korrektur von Inhomogenitäten in der Varianz der Messungen festgestellt werden. Validierung der korrigierten Messungen zeigt nur geringe Änderungen im Hinblick auf (relative) Fehlermetriken. Änderungen in den Bodenfeuchtetrends aufgrund der Korrektur werden in einigen Gebieten festgestellt. Die Korrektur von Inhomogenitäten verbessert global betrachtet die Qualität des Datensatzes, wobei der das Klima beschreibenden Anteil des Signals weitestgehend erhalten bleibt. Eine der drei Methodiken (“Quantile Category Matching”) wird gegenüber den anderen als bevorzugt identifiziert.

Acknowledgements

I want to thank Prof. Dorigo for supervising this work and for the opportunity to participate in the CLIMERS research group. I am especially thankful for his confidence in my work and the chance to contribute to the ESA CCI SM project. I want to thank all people from the Department of Geodesy and Geoinformation at TU Wien for everything they taught me during my studies. Special thanks go to my colleagues in the Remote Sensing research groups under Prof. Dorigo and Prof. Wagner for always being so dedicated, open and supportive. I want to especially thank Alexander Gruber (KU Leuven), Chun-Hsu Su (University of Melbourne) and Tracy Scanlon (TU Wien) for their feedback and many useful discussions during writing this thesis. I want to acknowledge all colleagues who have contributed to the success of ESA CCI SM in the past.

Contents

| | |
|---|-------------|
| Abstract | I |
| Kurzfassung | II |
| Acknowledgements | III |
| List of Figures | VI |
| List of Tables | VI |
| Acronyms | VIII |
| 1 Background | 1 |
| 1.1 Soil Moisture - Essential Climate Variable | 1 |
| 1.2 Satellite Remote Sensing | 2 |
| 1.2.1 Microwave Remote Sensing | 3 |
| 1.2.2 Passive MRS | 4 |
| 1.2.3 Active MRS | 6 |
| 1.2.4 Variable Retrieval | 7 |
| 2 Introduction | 10 |
| 3 Datasets and preprocessing | 12 |
| 3.1 ESA CCI SM | 12 |
| 3.2 MERRA2 | 13 |
| 3.3 ECMWF reanalysis | 15 |
| 4 Methods | 16 |
| 4.1 Break Detection | 16 |
| 4.2 Break Correction | 17 |
| 4.2.1 Procedure Overview | 17 |
| 4.2.2 Method 1: Linear Model Pair matching (LMP) | 22 |
| 4.2.3 Method 2: Higher Order Moments adjustment (HOM) | 24 |
| 4.2.4 Method 3: Quantile Category Matching adjustment (QCM) | 26 |
| 4.3 Evaluation | 28 |

| | |
|--|-----------|
| 5 Results and Discussion | 30 |
| 5.1 Global break detection and correction | 30 |
| 5.2 Analysis of intermediate results | 34 |
| 5.3 Impact of adjustment | 38 |
| 5.3.1 Longest Homogeneous Period | 38 |
| 5.3.2 Comparison against independent reanalysis data | 40 |
| 5.3.3 Trend analysis | 41 |
| 6 Conclusion | 44 |
| References | 46 |
| Supporting Information | |

List of Figures

| | | |
|----|---|----|
| 1 | Atmospheric Windows | 4 |
| 2 | ESA CCI SM v04.5 COMBINED Sensors and Sensor Periods | 12 |
| 3 | Period nomination for Break Detection and Correction | 20 |
| 4 | Process cycle | 21 |
| 5 | Linear Model Pair Adjustment | 23 |
| 6 | Higher Order Moments Adjustment | 24 |
| 7 | Higher Order Moments Residuals Fitting | 25 |
| 8 | Quantile Category Matching Adjustment | 27 |
| 9 | Global Break Detection Before and After Correction | 33 |
| 10 | Detailed Break Detection and Evaluation (1998-01-01) | 34 |
| 11 | Detailed Break Correction and Evaluation (1998-01-01) | 36 |
| 12 | Length of the Longest Homogeneous Period | 38 |
| 13 | Increase in Length of the Longest Homogeneous Period | 39 |
| 14 | Global Evaluation - ubRMSD | 41 |
| 15 | Change in Global Trends (QCM) | 42 |
| 16 | Global SM Trends | 43 |

List of Tables

| | | |
|---|--|----|
| 1 | ESA CCI SM Passive Sensors | 6 |
| 2 | ESA CCI SM Active Sensors | 7 |
| 3 | Break Detection Statistics | 32 |
| 4 | Break Corrections Statistics | 37 |

Acronyms

CCI Climate Change Initiative

CDF cumulative distribution function

CDR climate data record

CF cumulative frequency

ECMWF European Centre for Medium-Range Weather Forecasts

ECV Essential Climate Variable

ESA European Space Agency

FK Fligner-Killeen test

GCOS Global Climate Observing System

GLDAS Global Land Data Assimilation System

GMAO Global Modelling and Assimilation Office

HOM Higher Order Moments adjustment

IFOV instantaneous field of view

IR infrared

KS Kolmogorov-Smirnoff test

LEO low Earth orbit

LMP Linear Model Pair matching

LOWESS Locally weighted scatterplot smoothing

LPRM Land Parameter Retrieval Model

MERRA2 Modern-Era Retrospective analysis for Research and Applications 2

MRS microwave remote sensing

MSE mean-square error

MSP model sub-period

NASA National Aeronautics and Space Administration

NDVI Normalised Difference Vegetation Index

QC quantile category

QCM Quantile Category Matching adjustment

RADAR radio detection and ranging

- RFI** radio frequency interference
- RMSD** root-mean-square-difference
- RSS** residual sum of squares
- RTM** Relative Tests with Model reanalysis reference data

- SAR** Synthetic Aperture Radar
- SM** soil moisture
- SNR** Signal-to-Noise ratio
- SSP** sensor sub-period

- TC** Triple Collocation analysis

- ubRMSD** unbiased root-mean-square-difference

- WARP** Water Retrieval Package
- WK** Wilcoxon rank-sum test

1 Background

1.1 Soil Moisture - Essential Climate Variable

Soil moisture (SM) plays an important role in the global climate system, as it not only acts as an indicator, but also as a driver of weather and climate. It takes a key role in the global water, energy and carbon cycle. Vaporisation of soil water cools the surface and fuels precipitation. SM affects (surface) water infiltration and runoff and therefore droughts and floods [1], [2]. Soil water can limit vegetation growth and therefore affect carbon uptake. It is an important input for weather predictions and affects agricultural decisions (e.g. for irrigation) and is therefore of economic interest. SM is part of feedback cycles that drive climate change [3].

Predictions on climate change and decisions based on potential future scenarios rely on understanding the environmental processes and interactions in the real world. SM acts as an important parameter in Earth system models, that are used in biogeoscience to model interactions between the environment and living organisms in the real world [3]. Reliable measurements allow scientists to analyse these interactions. Global measurements are necessary due to the complexity of the Earth system, and even though in situ measurements of climate variables (temperature, precipitation, soil moisture, ...) are considered to be the observation type of highest quality, they cannot provide the needed global coverage [4]. In addition, long-time measurements are needed to detect changes over time. The Climate Change Initiative (CCI) of the European Space Agency (ESA) is a program to provide scientists with these measurements. Therefore global, long-term satellite measurements for multiple Essential Climate Variables (ECVs) are collected and homogenised within the program [5]. SM is classified as an essential climate variable by the Global Climate Observing System (GCOS) [6] and ESA CCI SM is the corresponding global, long-term soil moisture data set that is generated.

Measurements of SM from space-borne sensor systems, that measure radiation emitted and/or reflected from Earth's surface in the microwave domain, have proved to be a valuable input for regional and especially large-scale (global) analyses and models [7], [8]. In situ SM observations continue to gain in importance as they are not only used directly within local hydrological studies, but also for calibration and validation of SM estimates from satellite observations and models [9], [10].

1.2 Satellite Remote Sensing

Compared to other remote sensing platforms (e.g. airborne systems), satellites can measure large areas with a (constant) high temporal coverage and are therefore especially suitable for monitoring large-scale processes in the atmosphere (e.g. meteorological measurements) [11] and on the ground (e.g. geophysical measurements and land cover classification). In return they generally provide data with lower spatial resolution compared to airborne systems (depending on the type of observations).

A large number of satellite platforms and an even larger number of sensors that are carried by them have been available over the last decades. Satellite sensors measure radiation that comes from Earth, from objects on its surface and from the atmosphere above in form of electromagnetic waves. Different sensors measure the energy carried by waves in various domains of the electromagnetic spectrum (frequency bands). Important spectral domains for remote sensing are the visible (optical remote sensing), infrared (IR) and the microwave domain.

Globally operating Earth observation satellites used in the ESA CCI SM dataset move on polar, sun-synchronous low Earth orbits (LEOs). Sun-synchronous orbits allow the satellite to pass the equator at the same time of day and night (depending on the orbit, referred to as the "ascending" or "descending" overpass) at different locations regularly. This orbit is especially important for systems that are affected by variations in the illumination from sun [12]. Notably platforms on geostationary orbits exist, where the satellite movement is in line with the rotation of Earth. They are not used in the ESA CCI SM dataset but - if available - could be used to e.g. improve the data quality/density in areas where SM from such sensors is available in future. The frequency at which a point on Earth is actually measured (the temporal resolution) depends on the orbit of the satellite, but also on the measurement setup of the sensor. Some sensors are even steerable and can therefore increase the number of times a point on the surface is observed when necessary (at the cost of observations at other locations). While the temporal resolution depends mostly on the platform that carries a sensor, the sensors themselves can be classified by the type of information they acquire. They vary in measurement principle and measured quantity and have different advantages, disadvantages and are therefore used for different purposes. Sensors can be specialised to one category or be a hybrid of two but trade-offs must be made in terms of the data resolution for categories that the sensor is not designed for [13].

1. **Spatial Information** : Imagers, high spatial resolution
2. **Spectral Information** : Spectrometers, high number of spectral bands.

3. Intensity Information : Radiometers/Scatterometers, high radiometric resolution.

Sensors measuring at different wavelengths and different characteristics of electromagnetic waves (e.g. amplitude, phase, polarisation) allow deriving more information on the observed scene and different (environmental) variables. E.g. spectral information is key when measuring gas contents in the atmosphere and intensity information (backscatter, brightness temperature) is affected by the dielectric properties (water content) of an object and the basis to derive SM [14]. The combination of multiple different sensors allows a more complete picture of the surface conditions on Earth.

Advanced techniques allow to overcome the physical limitations of sensors to a certain degree. E.g. Synthetic Aperture Radars (SARs) make use of the platform movement and coherently recorded signals to identify phase shifts within (sub)regions of the antenna beams to reach much higher spatial resolutions than "normal" radars (therefore simulating a physically much larger antenna). Other interferometric techniques use differences in the measurements of multiple (real or synthetic) antennas to enhance the single measurements, e.g. by reducing antenna noise using signal-cross-correlations in the passive case (e.g. SMOS).

1.2.1 Microwave Remote Sensing

Microwaves are short compared to radiowaves (hence their name) and long compared to visible light. Microwaves in the SM-relevant domain (for wavelengths $\lambda \approx 3\text{-}30$ cm, and frequencies $f \approx 1\text{-}10$ GHz) are mostly unaffected by atmosphere, rain and clouds as indicated by the atmospheric windows in Fig. 1¹. They can therefore be used to measure radiation from the surface of the Earth under nearly all atmospheric conditions. Microwaves carry comparably low amounts of energy, measurements are therefore usually noise-like in appearance. Shorter waves (including the visible domain) on the other hand carry more energy, but are affected not only by the surface but also by the atmosphere (and its composition) and therefore can be used to monitor e.g. gas concentrations or cloud cover. The following frequency bands are of interest in terms of SM retrieval (after "IEEE Standard Letter Designations for Radar-Frequency Bands" [15]): L (1-2 GHz, 15-30 cm), S² (2-4 GHz, 7.5-15 cm), C (4-8 GHz, 3.9-7.5 cm), X (8-12.5 GHz, 2.4-3.8 cm), Ku, K and Ka (12.5-40 GHz, 0.75-2.4 cm).

Radiation in the described domains is either emitted by the Earth directly (measured

¹Original Image: NASA public domain, acquired from Wikimedia Commons

²Not in ESA CCI SM

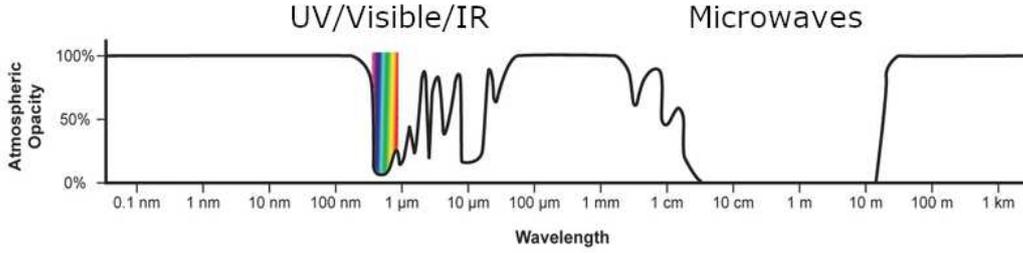


Figure 1: Atmospheric opacity of electromagnetic waves for different wavelengths.

in the passive case) or (back)scattered radiation after irradiation from the measuring sensor itself (active case). In any case sensors do not rely on direct illumination from the sun, in contrast to optical sensors, which measures visible light (which is emitted by the sun and that our eyes are therefore sensitive to). Other than light the longer microwaves can to a certain degree penetrate through vegetation and soil layers, depending on the composition.

The ESA CCI SM dataset combines SM measurements that are derived from the measurements of radiometers and scatterometers, i.e. from active and passive microwave remote sensing (MRS) systems. Therefore the basic measurement principles are described in the next sections.

1.2.2 Passive MRS

Microwave radiometry measures the thermal radiation that is emitted by any object with kinetic temperature $> 0^\circ K$. At a temperature of $\approx 300^\circ K$ Earth emits phase incoherent long and short wave radiation. The largest part is emitted in the IR domain ($\approx 10 \mu m$), while passive MRS measures the intensity of the radiation that is emitted at $\approx 1-20$ GHz.

Planck's Law describes the electromagnetic radiance or spectral brightness (emitted power) of a black body under a certain temperature and with a certain frequency. For the microwave domain at the Earth's surface temperature, the Rayleigh-Jeans law is a simple (linear) but good approximation to Planck's Law and can be used to describe spectral brightness L_f under a certain brightness temperature (T_B) as in (1).

$$L_f = \frac{2f^2 k T_B}{c^2}, L_\lambda = \frac{2k T_B}{\lambda^2} \quad (1)$$

As objects in reality are no blackbodies, a difference between the brightness temperature (T_B) and the (actual) kinetic temperature (T), depending on the emissivity (ϵ), the

ability to emit radiation) of an object is evident as described in (2) and (3).

$$L_{f,greycbody}(T) = L_{f,blackbody}(T_B) = \epsilon L_{f,blackbody}(T) \quad (2)$$

and

$$T_B = \epsilon T \quad (3)$$

This is the basis that allows passive microwave remote sensing to retrieve parameters related to the emissivity of an object by measuring the emitted radiation from Earth [16]. Water has a high emissivity (0.96) compared to dry soil and radiometers can be used to derive SM.

Measurements of the target brightness temperature of the surface are disturbed by atmospheric emissions and interference, e.g. radio frequency interference (RFI) [17] and by emissions from the radiometer itself, leading to a low Signal-to-Noise ratio (SNR) for passive measurements. Due to the low energy amounts that are emitted by Earth in the measured domain, a large instantaneous field of view (IFOV) is therefore necessary to increase the collected energy by the sensor. This results in a relatively low spatial resolution of the collected information.

Long-wave radiation emitted by Earth is prior absorbed (short-wave) radiation from the sun. As shown in (4), the amount of radiation that is absorbed (a) is equal to the amount that is emitted (ϵ). The part of the radiation that is not absorbed, is reflected (r).

$$\text{with } r + a = 1 \quad \text{and} \quad a = \epsilon \quad \rightarrow r = 1 - \epsilon \quad (4)$$

Eq. (4) indicates the link between passive and active MRS (described in more detail by Kirchhoff's Law). In the active case the surface is illuminated artificially and the reflected radiation (that depends on the surface emissivity) is the quantity of interest.

Passive sensors combined in the ESA CCI SM dataset and their characteristics are described in Table 1 (adapted from Gruber *et al.* [18]):

Table 1
PASSIVE SENSORS IN ESA CCI SM V04.5 COMBINED

| Sensor | SMMR | SMM/I | TMI ¹ | AMSR-E | WindSat | SMOS | AMSR2 |
|---------------------------|------------|------------|------------------|------------|------------|------------|------------|
| Platform | Nimbus 7 | DMSF | TRMM | Aqua | Coriolis | SMOS | GCOM-W1 |
| Band | C | K | X | C/X | C/X | L | C/X |
| Spatial Res. ² | 150 × 150 | 69 × 43 | 59 × 36 | 76 × 44 | 25 × 35 | 40 × 40 | 35 × 62 |
| Period from | Jan 1979 - | Sep 1987 - | Jan 1998 - | Jul 2002 - | Oct 2007 - | Jan 2010 - | Jul 2012 - |
| to | Aug 1987 | Dec 2007 | Dec 2013 | Oct 2011 | Jul 2012 | Dec 2018 | Dec 2018 |

¹ between 37° N and 37° S ² in km^2

1.2.3 Active MRS

Active microwave remote sensing, or radio detection and ranging (RADAR), measures radiation in form of pulses that are scattered back from the surface after actively irradiating the ground (receiving and transmitting unit is the same). Pulse measurements (frequency modulation is used in modern chirp radars to improve the radiometric resolution) allow precise measurement of the distance between the transmitter and the scatterer. Scattering describes the reflection and diffusion of radiation through the interaction with particles. For example, atmospheric scattering of sunlight changes the observed colour of the sky over the day. For active MRS the scattering interaction happens mainly at the surface as well as in the ground, allowing to subsequently derive geophysical parameters from the received (backscattered) radiation.

This is described in the radar equation (5), which links the received power (P_r) at the satellite sensor, with respect to the transmitted power (P_t), antenna properties - gain (G_t) and effective area (A_r) - and the distance between target and (moving) transmitter/receiver (R_t , R_r) as well as the radar cross section (σ) of the the target [19].

$$P_r = \frac{P_t G_t A_r \sigma}{(4\pi)^2 R_t^2 R_r^2} \quad (5)$$

and

$$\sigma = A_{eff}(1 - f_a)G_s \quad (6)$$

σ (in m^2) depends on shape and dielectric properties of a scatterer and on the incident radiation (frequency, polarisation, angle). A_{eff} is the effective area, irradiated on the ground, $(1 - f_a)$ is the radiation that is *not* absorbed and G_s is the additional contribution of the target (gain) to the received signal.

In Earth observation, σ is a complex combination of contributions from multiple

sources due to the inhomogeneous nature of Earth’s surface and ground. Normalised by the area - therefore in the unit of $[m^2m^{-2}]$, often expressed on a logarithmic scale in dB - the radar cross section is named the backscatter coefficient (σ^0). It is the quantity used in retrieval models to derive geophysical variables related to the geometry/ roughness and the dielectric properties (affected mainly by water content) of the measured area, similar as described for the passive case.

Scatterometer observations in ESA CCI SM have similar spatial resolutions as the passive products, which fits to the fact that the data set aims to provide global observations mainly for climate studies. SM from imaging radars (SAR) with higher spatial resolution is becoming available [20] and could be utilised in the product generation in the future.

The characteristics of active sensors that are combined in the ESA CCI SM dataset are described in Table 2 (adapted from Gruber *et al.* [18]):

Table 2
ACTIVE SENSORS IN ESA CCI SM V04.5 COMBINED

| Sensor | AMI-WS | AMI-WS | ASCAT-A | ASCAT-B |
|---------------------------|------------------------|------------------------|------------------------|------------------------|
| Platform | ERS-1/2 | ERS-2 | MetOp-A | MetOp-B |
| Band | C | C | C | C |
| Spatial Res. ¹ | 50×50 | 25×25 | 25×25 | 25×25 |
| Period from to | Jul 1991 - Dec 2006 | May 1997 - Feb 2007 | Jan 2007 - Jun 2018 | Nov 2012 - Jun 2018 |

¹ in km^2

1.2.4 Variable Retrieval

Satellite measurements contain contributions of multiple objects and processes in the observed scene in terms of the measured radiation. Disentangling the single contributors and retrieving different (in the case of soil moisture geophysical) variables with respect to the observed scene is the main challenge when exploiting Earth observation sensor measurements. The two main steps of parameter retrieval are the creation of a forward model and the subsequent inversion of the model [21].

The forward model describes the observable quantity (radiation that is measured by the satellite) for a certain observation scene and under certain measurement conditions. It therefore takes into account parameters for one or multiple (atmospheric or geophysical) variables of interest, for the (known) measurement configuration (e.g. observation time, frequency, etc.) and for the state of the observed quantity (that is measured in

practice). The number of parameters that are considered defines the complexity of the model. In the case of SM, surface radiation is not only affected by the soil properties, but also by water in the vegetation above. Volume models are therefore preferred over bare soil models in most cases in practice. They allow (to a certain degree) differentiation between the (backscatter) contribution from vegetation and from soil and therefore retrieving vegetation variables such as vegetation water content. Still, dense vegetation is problematic when deriving SM and therefore areas with dense vegetation coverage (e.g. tropical forests) are masked out in ESA CCI SM. We differ between three kinds of forward models:

1. **Theoretical models** : Aim to describe the state of the observed quantity using combinations of methods and theorems with researched (theoretical) physical considerations directly. No ancillary data is necessary.
2. **Empirical models** : Aim to describe the state of the observed quantity using a priori collected data and analyses (possibly from an experimental, supervised setup). Together with reference data of the parameter(s) of interest, the model is tuned to perform under various conditions.
3. **Semi-empirical ("hybrid") model** : Aim to describe the state of the observed quantity using a hybrid (theoretical and empirical) approach. The empirical function is fitted to theoretical results of the parameter(s) of interest and tuned using reference observations.

Inversion of the forward model then allows deriving the searched (unknown) geophysical parameters by using the actual measurements as input. Model inversion is an ill-posed problem because satellite measurements cover large areas with a multitude of (dependent) geophysical/atmospheric parameters and because the number and of available independent observations is limited. This means that the same (or a very similar) result can be found from a (large) number of different combinations of model parameters. Model calibration is necessary to tune parameters of a model so that the inversion returns the optimal description of a geophysical variable in reality. Reference measurements (station measurements on the ground) are important for calibration and subsequent (independent) verification of the derived parameters. Different approaches for minimising the discrepancies between the predicted and observed parameters at reference locations are available. Quantitative methods aim to find a set of model parameters to optimise the model outputs in terms of various performance metrics [22]. Still, the

derived output will never represent the true state of a variable, due to the unavoidable errors and limitations of the measurement sensor (spatial, spectral, temporal and radiometric resolution) and the necessary generalisations and simplifications that are made when creating the model. This has to be considered when using the data.

In the ESA CCI SM dataset, volumetric SM (in m^3m^{-3}) from all passive sensors is derived using the Land Parameter Retrieval Model (LPRM) [23], [24]. Soil moisture from scatterometer backscatter measurements is derived using the Water Retrieval Package (WARP) [25], [26] (based on the TU-Wien Method [27]) for all active sensors.

The following chapters are also found in Preimesberger *et al.* [28].

2 Introduction

As part of ESA’s Climate Change Initiative, global long-term SM data sets are derived from space-borne active and passive microwave remote sensing instruments [18], [29]–[31]. Due to the limited operational period of single satellite systems, ESA CCI SM uses multiple SM products from ongoing and past remote sensing programs to create harmonised, merged, long-term data sets. As multiple different constellations of SM products from microwave sensors with varying properties and characteristics (active and passive) are combined, inhomogeneities (or structural breaks) at the transitions between the merged sensor periods (Fig. 2) may not be completely removed by the merging process. Structural breaks in climate data records can negatively influence the conclusion of scientific analyses, as they make data unrepresentative of the actual climate variability over time [32], [33]. Therefore especially for trend detection, inhomogeneities may conceal the true characteristics of the measurements in use and may lead to misinterpretation of observed phenomena.

Detection and homogenisation of structural breaks has been subject of many studies on terrestrial observations, mainly on measurements of temperature and precipitation. Potential reasons for inhomogeneities in climate records were found due to urbanisation, station movement, or change in measurement sensor type [34]–[36]. Independent of the measured variable, studies that test the homogeneity of climate records suggest that no long-term measurement data set is free of inhomogeneities.

More and more data from Earth observation satellites is becoming available and an increasing number of disciplines is using it. Available data records are combined to overcome the temporal limitations of single sensor products [5] and therefore the detection of inhomogeneities in these records is becoming the topic of research (e.g. Pinzon *et al.* [37] on Normalised Difference Vegetation Index (NDVI), Brinckmann *et al.* [38] for solar irradiance data, Schroeder *et al.* [39] for satellite water vapour). However, correcting detected breaks in combined satellite data sets is still a major challenge. SM records in general have received little to no consideration in this regard. This is most likely because most of them have covered relatively short time periods until now. Methods for detection and adjustment depend on the type of climate record. Temperature and SM are continuous variables, whereas for example precipitation is discontinuous [40]. SM is limited by the soil saturation. Correction methods may also depend on the type of shifts that are found in the data set, e.g. breaks in mean and variance. Some studies

also adapt correction methods to regional circumstances, to better take into account - for example - local climate or weather [41], [42].

In this study we compare three different homogenisation methods. They aim to adjust inhomogeneities in mean and variance in the global ESA CCI SM v04.5 COMBINED multi-satellite product [18], [29]. The goal is to match homogeneous sub-periods that were not sufficiently homogenised during the initial merging process. We aim to reduce the number of detected breaks and optimise the adjustment process in terms of minimising error metrics with respect to (mostly) independent reference datasets. In Section 3 the candidate (ESA CCI SM) and reference data used in the homogenisation and evaluation process are described in more detail. The methods for the detection of breaks are shortly reviewed in Section 4. Section 4 also looks at developing, adapting and implementing three separate break correction methods within a satellite SM specific framework. It further contains short descriptions of the common methodology to evaluate the corrections. Section 5 compares the results for different methods in terms of the reduction of breaks in the ESA CCI SM v04.5 COMBINED data set. The impact of adjustment on error metrics and long term trends in the data set is also described there.

3 Datasets and preprocessing

3.1 ESA CCI SM

ESA CCI SM is a long term climate data record (CDR) on global SM, based on multiple satellite remote sensing products (as indicated by sensor names in Fig. 2 for the ESA CCI SM v04.5 COMBINED product). SM measurements from an (increasing) number of satellite-based active (scatterometers, Fig. 2 blue names) and passive (radiometer, Fig. 2 red names) microwave observation systems are used in the merging process. ESA CCI SM is available as three separate products, which contain only active, only passive and both types of measurements (the here used COMBINED product) [18], [29]. Weights for merging available sensor products are derived from Signal-to-Noise ratio (SNR) estimates from Triple Collocation analysis (TC) [31]. Homogenisation of the single products is achieved via scaling by matching cumulative distribution functions (CDFs) with respect to SM simulations from the Global Land Data Assimilation System (GLDAS) Noah v2.1 [43] land surface model as the common reference data set. The most recent version of ESA CCI SM spans over a period of over 40 years (1978-10-26 to 2018-12-31).

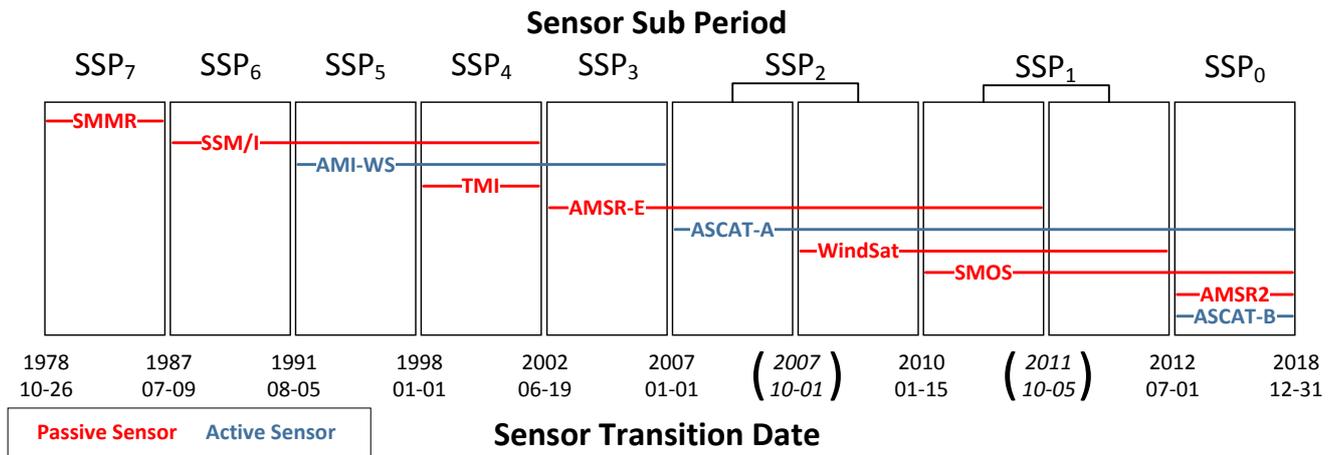


Figure 2: ESA CCI SM v04.5 COMBINED sensor periods: Horizontal lines describe the temporal coverage of active (red) and passive (blue) sensor SM products. Boxes indicate changes in the set of merged sensor products. Sensor transition dates accord to the dates of potential breaks in the data set and define the sensor sub-periods (SSPs). SSP_0 to SSP_7 therefore describe the minimal assumed homogeneous sub-periods in the ESA CCI SM product. Two transitions are excluded in this study (parenthesised dates) to avoid sub-periods that are shorter than one year.

ESA CCI SM COMBINED observations are representative of water in the top layer of the soil (first few centimetres, depending on the radiometric properties of the merged sensors at any given time) and expressed in volumetric units (m^3m^{-3}). The data is organised on a global quarter degree grid. Observation flags as well as error estimates for single observations are provided with the SM measurements. We use only the unflagged ("good") observations of ESA CCI SM in this study. The original ESA CCI SM daily images were converted into a time series format with varying temporal coverage, depending on the number and orbits of the sensor platforms in use. From the nine sensor transition dates as shown in Fig. 2 (excluding the start and end date), seven are considered in this study, excluding October 2011 and October 2007 to avoid sensor sub-periods that are shorter than one year. This results in a total of eight (assumed) homogeneous sensor sub-periods (SSPs) in between the potential breaks. The ESA CCI SM product is released together with a static mask of densely vegetated areas, which covers about 6% of all global land points (excluding Antarctica). SM retrievals from satellite observations are unreliable in these (tropical rainforest) areas [44]. Therefore they are excluded in all results and areas are marked in most maps with dark green colour.

3.2 MERRA2

In this study we use *relative* methods to detect and correct breaks. This means that for testing and adjusting the homogeneity of single candidate observation series we rely on (homogeneous) reference observations for comparison in non-parametric statistical tests [45] and also for characterisation of detected breaks for subsequent adjustment. The candidate and reference datasets must be similarly influenced by climate and weather to distinguish between artificial and natural shifts. Assuming that the reference is homogeneous, artificial breaks (that are introduced by a change in the measurement setup) would only appear in the candidate. Naturally caused changes would be represented in both datasets [46]. Finding a suitable reference for a candidate is a fundamental part of all relative break testing/adjustment methods. Most studies prefer using in situ measurements to analyse the homogeneity of other nearby stations over reanalysis data as they are considered to contain less (and better documented) breaks [47]. In situ SM observation sites are, however, limited in coverage and generally span shorter time periods than satellite observations [4]. Therefore, the use of in situ reference data for homogenising (multi-)satellite observations is not feasible in our case and a global long-term reference product is needed. Land surface models replicate the interactions of

climate processes on Earth and therefore assimilate observations that indirectly affect the simulated SM state (e.g. precipitation). Satellite SM observations are not directly assimilated in most models, which are therefore as independent as possible from the ESA CCI SM data set. The use of reanalysis reference data for detection and adjustment of breaks in climate variables is a common practice in homogenisation studies for cases where in situ measurements are not sufficiently available (e.g. Azorin-Molina *et al.* [48] on wind speed observations). A single series from the reference data set can be used directly - the nearest neighbour is usually considered to show the most similar naturally caused deviations - or multiple nearby references can be combined to create a benchmark for the candidate [49]. As spatial inconsistencies in model SM simulations are assumed to be small, the creation of merged reference series (e.g. from neighbouring pixels) is considered not to be necessary within this study.

In this study, we used the National Aeronautics and Space Administration (NASA) Global Modelling and Assimilation Office (GMAO) Modern-Era Retrospective analysis for Research and Applications 2 (MERRA2) data set [50], [51]. MERRA2 provides hourly soil water simulations starting with the beginning of the year 1980 until present. It therefore covers a similar time period as ESA CCI SM. We select SM simulations from the (single) nearest grid point in the MERRA2 reanalysis reference data set to act as a the reference for each according candidate. The data is computed on a global $0.5^\circ \times 0.625^\circ$ grid. The original hourly fields were temporally resampled to daily averages to match the temporal sampling of the ESA CCI SM product.

By using MERRA2 SM, we follow Su *et al.* [45], who found SM from the MERRA-Land model [52] skillful in terms of global detection of inhomogeneities. Both MERRA2 and MERRA-Land use observation-based precipitation forcing for creating SM simulations [53]. Model SM is therefore assumed to contain less inhomogeneities compared to ESA CCI SM, especially in areas with precipitation measurements. We use the "Water Surface Layer" (referred to by the official short name SFMC) variable of MERRA2 (in m^3m^{-3}). It is representative of SM in the top ten centimetres of the soil and therefore comparable to ESA CCI SM observations. A bias between the two is expected, because of the different model scales and computational differences of MERRA2 compared to the scaling reference used in the production of ESA CCI SM (the GLDAS Noah v2.1). This bias is accounted for by matching the CDF of the reference series to that of the candidate series at each grid cell [54], [55]. A predefined set of percentiles (0, 5, 10, 30, 50, 70, 90, 95, 100) with a linear interpolation between them is used.

3.3 ECMWF reanalysis

We use two reanalysis products from the European Centre for Medium-Range Weather Forecasts (ECMWF) for validation [56], i.e. simulations from the recently released ERA5 reanalysis [57], which is the successor to the ERA-Interim reanalysis [58], as well as from the according ERA-Interim/Land product [59]. ERA5 provides global, sub-daily simulations of variables for land, atmosphere and ocean waves. 6-hourly images (starting at 0:00 UTC) with a spatial resolution of $0.25^\circ \times 0.25^\circ$ of the topmost (0-7 cm) of four available soil water layers - the "Volumetric soil water layer 1" (SWVL1) variable between 1979 and 2018 - were temporally resampled (arithmetic mean of 6-hourly simulations) to match the daily resolution of ESA CCI SM. ERA-Interim/Land uses near-surface meteorological fields of ERA-Interim together with precipitation adjustments to force the HTESSEL land surface model, leading to improvements in terms of land water simulations compared to the ERA-Interim reanalysis. A temporally extended version of the data set was used, in which the "Volumetric Soil Moisture of Layer 1" variable is available in the period from 1979 until 2015 with a horizontal resolution of ~ 80 km. Similar as for ERA5, ERA-Interim/Land simulations were temporally resampled to daily averages.

4 Methods

4.1 Break Detection

To detect breaks in ESA CCI SM v04.5 COMBINED, we follow Su *et al.* [45]. This method uses Relative Tests with Model reanalysis reference data (RTM) on SM difference series (ΔSM) between candidate observations from ESA CCI SM and the corresponding reference simulations from MERRA2. Difference series do not include naturally caused variations (that are not breaks) between SSPs - since they should exist in both datasets. The null hypothesis for break detection therefore is, that there are no significant differences in the first two moments of distributions of ΔSM before and after the tested sensor transition date. In the ESA CCI SM dataset the potential break dates are defined by variations in the sensor products that are combined (the SSPs shown in Fig. 2) and therefore known beforehand. SSPs form the minimum input sets for break detection and correction. As will be shown later, input data for testing and correction are not restricted to SSPs but can be extended across negatively tested (i.e. homogeneous) transition dates. We formulate pairs of adjacent SSP difference series as described in (7) and (8) as the input sets for statistical testing.

$$\Delta SM_{SSP_{i,0}} = CAN_{SSP_{i,0}} - REF_{SSP_{i,0}} \quad (7)$$

$$\Delta SM_{SSP_{i,1}} = CAN_{SSP_{i,1}} - REF_{SSP_{i,1}} \quad (8)$$

For each sub-period $SSP_{i,j}$, $i \in [0, \dots, 6]$ is the identifier of the tested transition date, with $i = 0$ as the latest date of potentially introducing a break. j marks whether the respective period is before ($j = 1$) or after ($j = 0$) the tested transition date that separates the two sub-periods (compare to Fig. 3b). For testing, common daily candidate (CAN) and reference (REF) values were jointly temporally resampled as monthly means with a required daily coverage per month of $> 30\%$ [45]. Therefore > 9 days per month are required to calculate the arithmetic mean that is then representative for the month, otherwise no value is assigned. Monthly averages are chosen as the input to reduce noise and outliers in each SSP, which could make detection of mean and variance breaks unreliable [45].

Breaks are detected by relative, nonparametric, statistical testing [32], [60], following two methods that were found applicable to the ESA CCI SM v02.2 COMBINED datasets by Su *et al.* [45]. First we use the nonparametric Wilcoxon rank-sum test (WK) [61] for detection of shifts in ranked, tie-corrected [62] difference values, which corresponds to an

assumed mean break in the candidate. Second the nonparametric Fligner-Killeen test (FK) [63] detects relative shifts in the variance of two differences series. Su *et al.* [45] found both methods skilful in terms of odds ratio (incorrect versus correct detection of an inhomogeneity) and Hanssen-Kuipers discriminant (skill of separating detected inhomogeneities from homogeneities) when comparing results from RTM to relative testing with reference data from in situ measurements. The resampled observations are again scaled before calculating ΔSM within the SSP pairs using linear regression scaling to reduce common short term biases that may exist across the tested transition date. Relative biases (i.e. breaks that should be detected) in mean and variance between the two periods are not affected by this scaling choice.

The input observations used in break testing need to comply with two predefined statistical requirements. 1) The number of monthly values in $SSP_{i,0}$ and $SSP_{i,1}$ is sufficiently large ($N > 10$). 2) Spearman's R (R_{SP}) between the candidate and reference is greater than 0.5 and significant ($p < 0.05$) across the combined sub-periods (joined period $SSP_{i,0}$ with $SSP_{i,1}$). If any of these two preconditions fails, the respective transition date is not tested.

We identify three possible types of sensor transition dates in ESA CCI SM in each location. 1) *Homogeneous* (negatively tested) transition dates: Both null hypotheses (by WK and FK) cannot be rejected. 2) *Inhomogeneous* (positively tested) transition dates: A statistically significant break in mean, variance (or both) is detected. 3) *Untested* transition dates: Both tests are not applied because one (or both) of the prerequisites is not fulfilled.

4.2 Break Correction

4.2.1 Procedure Overview

To homogenise detected breaks in ESA CCI SM v04.5 COMBINED, an iterative adjustment approach is implemented, starting from the most recent SSP. For each candidate series we start by detecting breaks between SSPs as described in Section 4.1 and as schematically shown in Fig. 3b.

Then - based on these (initial) test results - breaks are quantified by comparing *extended* SSPs, which are referred to as model sub-periods (MSPs). This means that the sensor periods before and after detected breaks are expanded across negatively tested (homogeneous) sensor transition dates (compare e.g. $MSP_{1,1}$ in Fig. 3c) or successfully removed ones (during previous correction iterations, as indicated by the dashed horizontal line part of $MSP_{2,0}$ in Fig. 3c). Extension is not performed across detected breaks

and untestable dates (compare e.g. $MSP_{1,1}$ respectively $MSP_{1,0}$ in Fig. 3c). This allows to increase the number of input values for correction which makes it more robust. To ensure that the period extension does not affect the test result, break detection is repeated with values in the MSPs (instead of the SSPs) before attempting to correct any re-detected breaks. If the break is not detected again using the extended period, it is not corrected. As an additional requirement to the re-detection, correction is only attempted if the candidate and reference correlate sufficiently with Pearson's correlation coefficient $R_P > 0.3$. This is done on top of the previous evaluations for testing, to make prerequisites for correction even more restrictive (step I. in Fig. 4). R_P is checked for both MSPs around the break (for the LMP and QCM methods, described in Sections 4.2.2 and 4.2.4), respectively only for $MSP_{i,0}$ (for the HOM method, described in Section 4.2.3).

In addition to finding MSPs, the initial break detection is also used to define ASPs (see Fig. 3c) for each iteration of correction and *between detected breaks*. As not all sensor transition dates can be tested, it is necessary to separate the input values used for quantifying a break (MSPs) from values that corrections are applied to (ASP). On the one hand this is done to avoid that derived corrections are distorted by inhomogeneous observations, which could be the case if $MSP_{i,1}$ included untestable subperiods that are potentially inhomogeneous. On the other hand this avoids the accumulation of previously corrected shifts at earlier, untestable transition dates. Notably, $MSP_{i,1}$ and ASP_i are equal in cases where there is no untestable transition date between two detected breaks (e.g. $MSP_{1,1} = ASP_1$ in Fig. 3c). Applying correction factors to ASPs, that were found between MSPs in cases when $MSP_{i,1} \neq ASP_i$ therefore assumes that values that are in ASP_i but not in $MSP_{i,1}$ are shifted accordingly. Then not only are the MSPs matched, but also untested or homogeneous transitions would remain unchanged (as will be shown later this is not always the case). This correction procedure (step II a. and II b. in Fig. 4) is evaluated to some extent in a third iteration of break detection - after adjustment - again using the MSPs as inputs (step III. in Fig. 4). We evaluate: 1) Whether a break was completely removed and 2) whether the relative bias between ASP_i and $MSP_{i,0}$ (i.e. before and after the break) has decreased.

If one of the two checks after adjustment fails, the correction is reverted and the values from before adjusting the current subperiod are restored (step IV a. in Fig. 4). Equations (9)-(11) describe the second (bias) constraint between the adjusted (ADJ) and uncorrected (CAN) candidate with respect to the reference (REF) data in the according

periods.

$$\Delta BIAS_{i,adjusted} \leq \Delta BIAS_{i,unadjusted} \quad (9)$$

with

$$\Delta BIAS_{i,adjusted} = |BIAS(ADJ, REF)_{ASP_i} - BIAS(CAN, REF)_{MSP_{i,0}}| \quad (10)$$

and

$$\Delta BIAS_{i,unadjusted} = |BIAS(CAN, REF)_{ASP_i} - BIAS(CAN, REF)_{MSP_{i,0}}| \quad (11)$$

As indicated by (9)-(11) and Fig. 3c, we aim to relatively fit the period before a break to the homogeneous period after the break. The most recent homogeneous sub-period ($MSP_{0,0}$) is referred to as the base period, as iterative adjustment progresses towards the beginning of the candidate series. Values in the base period are never changed. The most recent homogeneous sub-period is chosen as the base period for the following reasons: First, because later periods of the ESA CCI SM v04.5 COMBINED dataset show higher temporal coverage compared to earlier periods due to the larger number of merged sensor products (compare to SI Fig. 5). Second, because the quality of used sensors has improved over time, resulting in the most accurate observations for later periods. Lastly, because compared to other sensor periods, with more than six years in length this is a relatively long (assumed) homogeneous period without changes in the merged sensor products.

Fig. 4 shows the complete previously described processing cycle for correcting multiple breaks at a single location and the three (interchangeable) principles (LMP, HOM and QCM in step II. of Fig. 4) to quantify the size of detected inhomogeneities and subsequently remove them. The goal is to identify one approach that not only reduces breaks in the ESA CCI SM data and improves its quality, but simultaneously preserves the (not break-related) characteristics.

To explain and compare the three different adjustment principles in an example case in the following sections, values for ESA CCI SM and MERRA2 in the homogeneous period between 2010-01-12 and 2018-12-31 for a single point in North Louisiana (Lat: 32.875° , Lon: -91.625°) are used. ESA CCI SM values before 2012-07-01 have been distorted by an arbitrary multiplicative factor of 1.1 and an additive bias of $+0.05 \text{ m}^3 \text{ m}^{-3}$ to create a synthetic mean-break of known size (detected by WK, also illustrated in SI Figs. 1, 2 and 3).

4.2 Break Correction

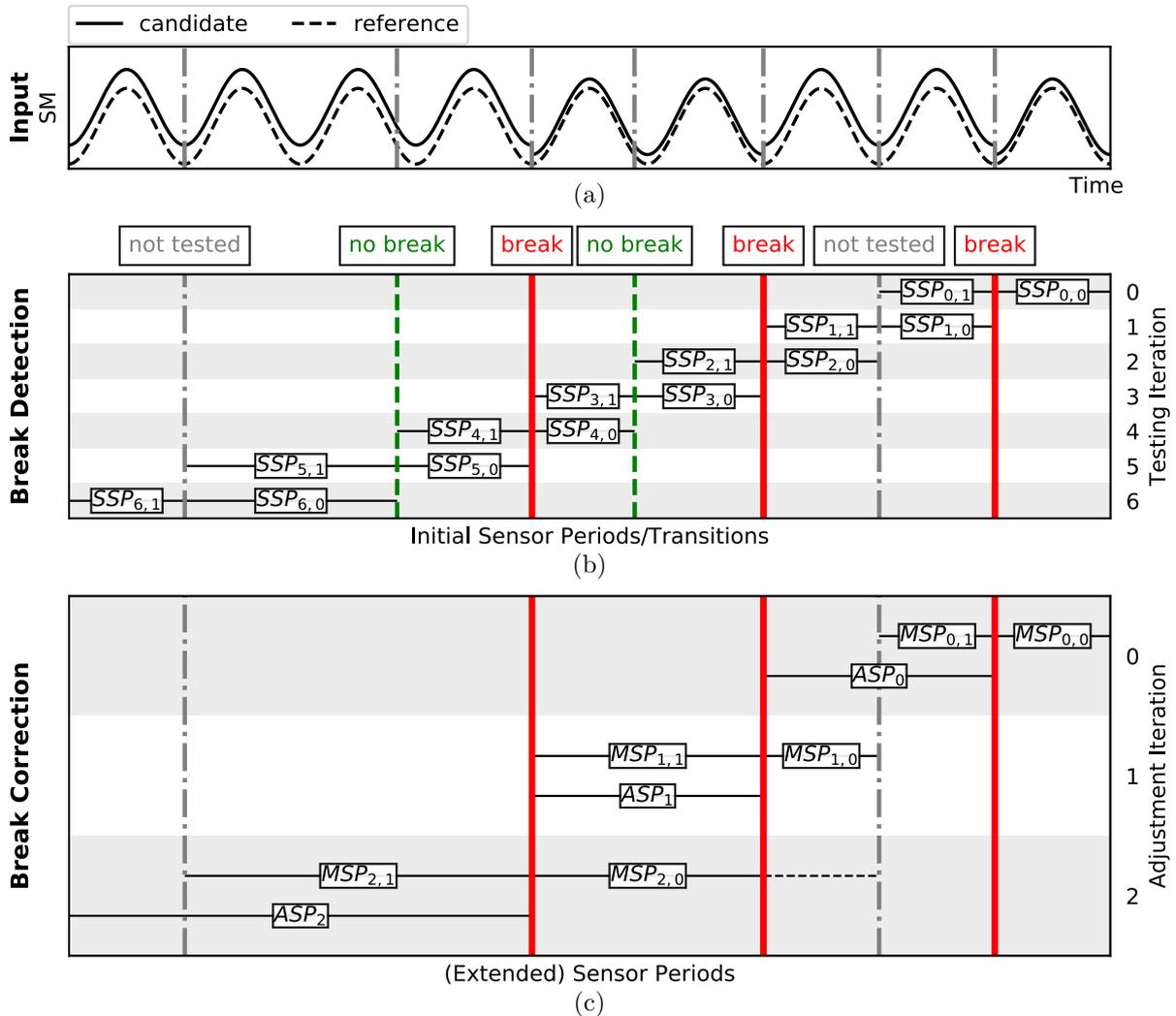


Figure 3: Schematic representation of period nomination for break detection (b) and correction (c) and concept of period extension for correction. The original sensor transition dates in each candidate series (vertical lines in the top segment) are initially tested as shown in the middle segment: Testing detects homogeneities (dashed green lines) and breaks (solid red lines) as well as un-testable transitions (dashed grey lines). The lower segment shows the period extension for adjustment based on the initial break detection, where MSPs for quantifying the break are extended across homogeneities (green) only. Due to untested transition dates, ASPs - periods that corrections derived from MSPs are applied to - cannot be equal to the respective MSPs and are therefore defined as the candidate observations between detected breaks.

4.2 Break Correction

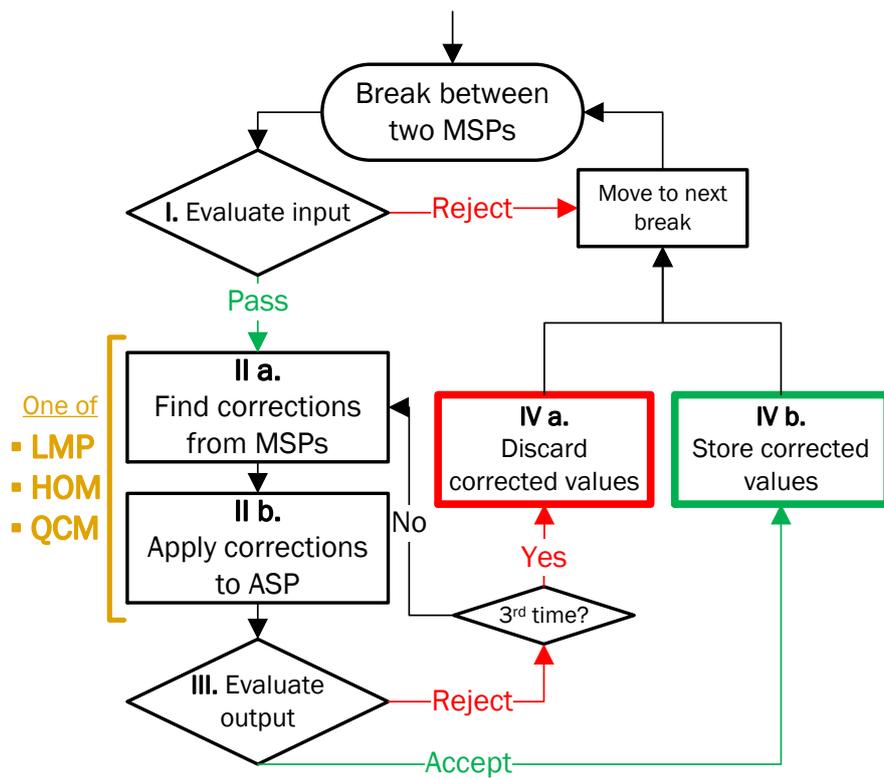


Figure 4: Process cycle of the iterative correction between detected breaks.

4.2.2 Method 1: Linear Model Pair matching (LMP)

For the first method we assume that detected relative breaks in each candidate can be quantified by differences in the parameters of two linear regression models [64]. The first model is found between the candidate and reference in $MSP_{i,0}$. The second one accordingly within $MSP_{i,1}$. Monthly resampled candidate and reference observations (as in Section 4.1) are used as inputs to model the linear relationship $Y = X \cdot [\alpha, \beta]^T + \epsilon$ with candidate observations (Y), the design matrix (X), the model intercept (α) and slope (β). ϵ describes the (minimised) residuals between the least-squares predicted and the observed values. A break (in mean or variance) in the candidate across $MSP_{i,0}$ and $MSP_{i,1}$ leads to differences in the model coefficients $\alpha_{MSP_{i,0}}$ and $\alpha_{MSP_{i,1}}$, respectively $\beta_{MSP_{i,0}}$ and $\beta_{MSP_{i,1}}$ of the two sub-periods. Discrepancies in α and β represent additive respectively multiplicative relative biases.

To reduce the impact of outliers on the models, in addition to using monthly values we exclude the 5% of observation pairs with the largest corresponding $|\Delta SM|$. It is assumed that single outliers are not caused by relative breaks between the compared periods as breaks affect the whole period.

Fig. 5 shows the linear regression models for the two MSPs around the introduced break in the example series in North Louisiana. The calculated model coefficients for the biased observations in $MSP_{1,0}$ were found with $\alpha = 0.15$ and $\beta = 0.75$ from 29 observations (Fig. 5a) and for $MSP_{0,0}$ with $\alpha = 0.10$ and $\beta = 0.68$ from 69 observations in this period (Fig. 5b). The introduced break is therefore mainly represented as a difference in the α coefficient in the example case.

As discrepancies in the model coefficients represent the size of the detected break they can be used to find required corrections as described in (12)-(14). CAN describes the monthly values, that are also used for finding the regression models. $CAN_{MSP_{i,1}}^{correction}$ contains monthly corrections for each candidate value before the introduction date of the detected break.

$$CAN_{MSP_{i,1}}^{correction} = (r \cdot CAN_{MSP_{i,1}} + c) - CAN_{MSP_{i,1}} \quad (12)$$

with

$$r = \left(\frac{\beta_{MSP_{i,0}}}{\beta_{MSP_{i,1}}} \right) \quad (13)$$

and

$$c = (\alpha_{MSP_{i,1}} - r \cdot \alpha_{MSP_{i,0}}) \quad (14)$$

To apply the corrections ($CAN_{MSP_{i,1}}^{correction}$) to observations in the extended period ASP_i ,

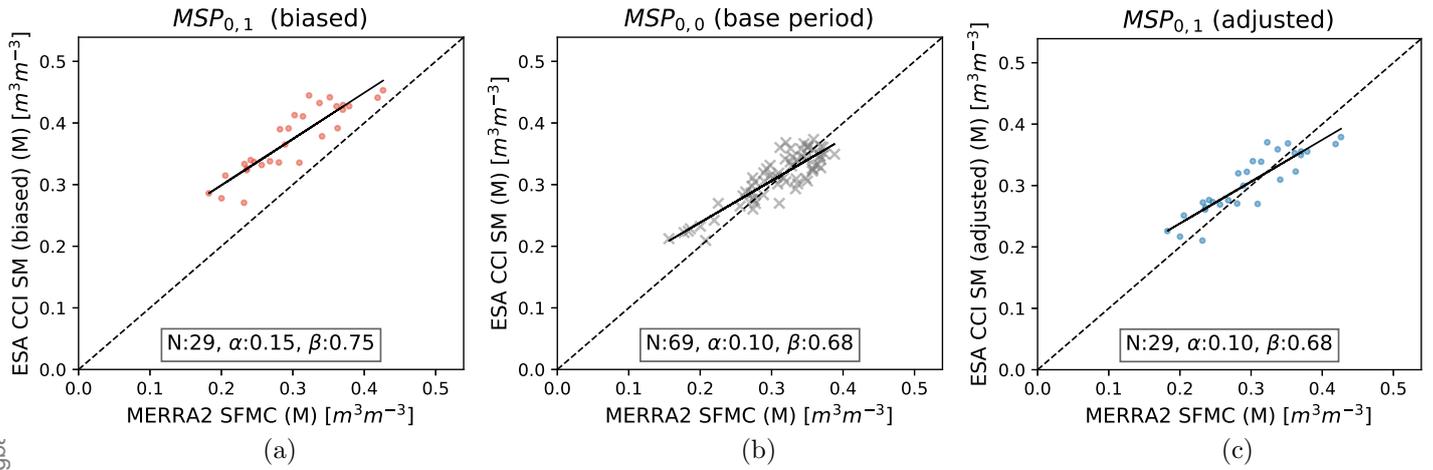


Figure 5: LMP adjustment - linear regression models for the two MSPs, i.e. before (a) and after (b) the detected break. The introduced bias leads to a discrepancy in the models coefficients. For comparison (c) shows the same period as (a), but after successful correction, which resulted in matching the model parameters to (b).

we follow Vincent *et al.* [42] and find one average correction for each month of $CAN_{MSP_{i,1}}^{correction}$ (i.e. the arithmetic mean of all corrections for January, for February, and so on). We then multiply the so found vector with the tridiagonal A-matrix (15) from Sheng *et al.* [65] to find a mid-month "target" value (at day 15) of each month. Linear interpolation between these "target" values leads to daily corrections that are - for each month - on average equal to the original monthly corrections.

$$A = \begin{pmatrix} 7/8 & 1/8 & & & \\ 1/8 & 6/8 & 1/8 & & \\ & \ddots & \ddots & \ddots & \\ & & 1/8 & 6/8 & 1/8 \\ & & & 1/8 & 7/8 \end{pmatrix} \quad (15)$$

These adjustments are then repeatedly applied (added) to the original inputs in the period to adjust (ASP_i), until the next break (or the beginning of the observation series). This way we adjust observations in $MSP_{i,1}$ and theoretically also match the remaining values in ASP_i accordingly without introducing (undetectable) relative differences between them.

4.2.3 Method 2: Higher Order Moments adjustment (HOM)

This homogenisation method was originally proposed by Della-Marta *et al.* [66] and successfully used to correct inhomogeneities in in situ temperature observations. It makes use of a quadratic regression model from the candidate and reference after a break to predict and correct observations before the break. Polynomial regressions allow considering potential nonlinearities, which may appear between satellite and reanalysis data, when quantifying breaks. They may be caused for example by (too) high reanalysis SM auto-correlation or varying discrepancies between model and satellite observations (e.g. for different seasons due to the unreliability of satellite SM estimates under frozen soil conditions).

We start by fitting a second order polynomial regression model between the ESA CCI SM candidate and MERRA2 reference series in the MSP after the break (a linear model is used in cases where $R_P > 0.8$ with $p < 0.05$ between the candidate and reference in $MSP_{i,0}$). Again monthly resampled and filtered values (as in Section 4.2.2) are used as inputs. The nonlinear model from candidate and reference time series for the base period ($MSP_{0,0}$) is shown in Fig. 6b.

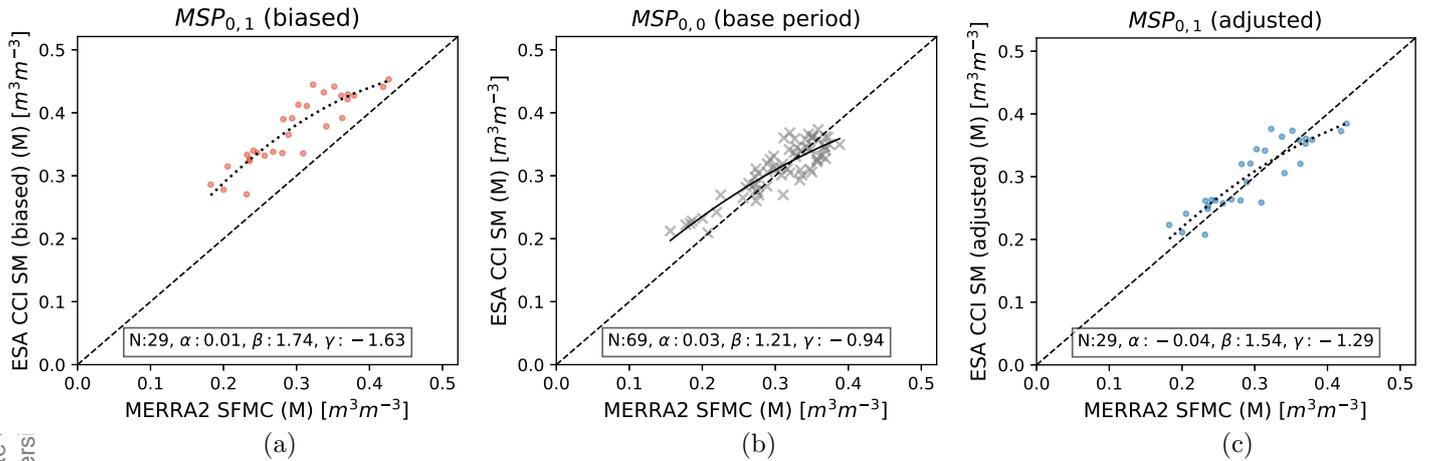


Figure 6: HOM adjustment - Second order polynomial regression model for the reference period of the same example series as in Section 4.2.2 from monthly resampled values after the introduced break (b). The model is used to create homogeneous predictions to find corrections for the biased candidate observations (a) and subsequently adjust them (c). The dotted lines in (a) and (c) indicate that the regressions are only shown for comparison and not used for correction.

Together with the reference data in $MSP_{i,1}$ the model is then used to create (homoge-

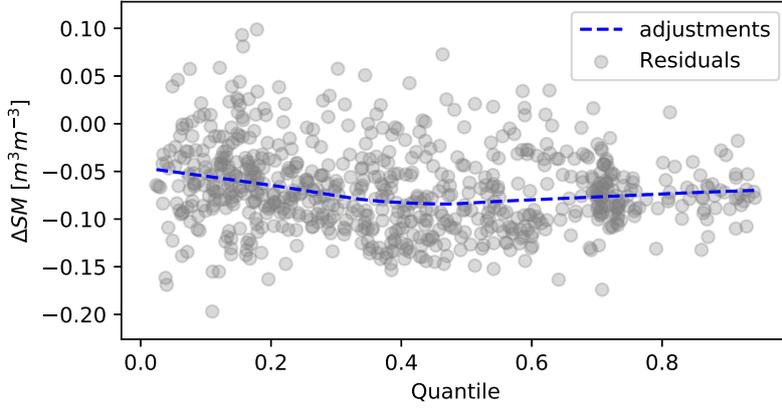


Figure 7: SM residuals between observed and predicted candidate (circles) and derived quantile adjustments (LOWESS fit with $\alpha = 0.6$, dashed blue line) within the period $MSP_{0,1}$ of the example time series.

neous) predictions of the candidate. We find adjustments for the candidate observations based on these predictions: We use a theory of L-moments approach [67] to fit either a Normal, a Pearson Type Three, a Generalised Normal or a Generalised Extreme Value distribution to the candidate in the MSP after the break (compare SI Fig. 2). A Kolmogorov-Smirnoff test (KS) [68] is used to compare the ensemble of possible theoretical CDFs to the candidate observations and predictions. The best fitting distribution is selected by minimising the maximum deviation (the KS statistics) between fitted theoretical and empirical CDF.

The so found CDF for the candidate after the break is used to calculate quantiles for the observations ($CAN_{MSP_{i,1}}$) and predictions ($PRED_{MSP_{i,1}}$) before the break, respectively the observations in ASP_i . We then fit a Locally weighted scatterplot smoothing (LOWESS) function [69] [70] to the SM differences (residuals) between the observations and predictions ($\Delta SM = CAN_{MSP_{i,1}} - PRED_{MSP_{i,1}}$) as shown in Fig. 7 for the demonstration case. LOWESS uses multiple linear estimates within local subsets of the residuals to fit a smooth polynomial. The factor α allows control over the smoothing by increasing/decreasing the subset size used for estimating the local regressions to combine. A fraction of $\alpha = 0.6$ was chosen to fit a curve, that varies only little and represents the distinct adjustment of each quantile. These corrections are then applied (added) to all candidate observations in ASP_i to relatively fit ESA CCI SM values before the break to observations in the MSP after the break with respect to changes in MERRA2.

4.2.4 Method 3: Quantile Category Matching adjustment (QCM)

This method was proposed by Wang *et al.* [71], who adjusted breaks in (de-trended) precipitation series (absolute method) and adapted by e.g. Vincent *et al.* [72] to adjust breaks in temperature series using nearby in situ reference observations (relative method). It builds on the assumption that adjustments for a detected break can be derived by detecting differences in a number of (median) empirical cumulative frequencies (ΔECF) of various quantile categories (QCs) between the candidate and the reference before and after the transition date. Each category contains observations over an equal range of quantiles. Differences are found by comparison of QC means of adjacent homogeneous sub-periods ($MSP_{i,0}$ and $MSP_{i,1}$). Relative differences between the candidate and the reference across the detected break then represent the size of the break for each QC separately.

We start with a maximum of four (quartile) categories because ESA CCI SM sensor periods can be short (compare Fig. 2) and include a limited range of theoretically possible observation values. Reasons for this can be for example low climate variability, geophysical restraints or inherited limits from the CCI scaling reference (GLDAS Noah) [29]. We find quantiles by sorting (ranking) data in ascending order by the candidate subset and find for each (unique) candidate observation the normalised cumulative frequency (CF) in both MSPs around a detected break. We divide the ranked observations of each MSP into the selected number of categories, so that each category spans over an equal range of CFs (again by candidate). For each $QC_i \in [1, \dots, QC]$ with $QC \leq 4$ in each MSP, we find the median CF of its according CF range and find the mean of the differences between the candidate and the reference in this category (ΔECF) as described in (16) and (17).

$$\Delta ECF \left(\frac{QC_i - 0.5}{QC} \right)_{MSP_{i,0}} = \text{mean}(CAN_{MSP_{i,0}, QC_i} - REF_{MSP_{i,0}, QC_i}) \quad (16)$$

$$\Delta ECF \left(\frac{QC_i - 0.5}{QC} \right)_{MSP_{i,1}} = \text{mean}(CAN_{MSP_{i,1}, QC_i} - REF_{MSP_{i,1}, QC_i}) \quad (17)$$

ΔECF for $MSP_{0,0}$ and $MSP_{0,1}$ for the example break are shown in Fig. 8a and 8b where the dots are the differences between the median CF of the candidate and reference. The median CF represents the frequency range of each category (indicated by the bars).

In cases where the number of categories is too high (i.e. one or more category does not

include any observations), we reduce QC by one until each category contains observations or only one category containing all observations for each MSP is left.

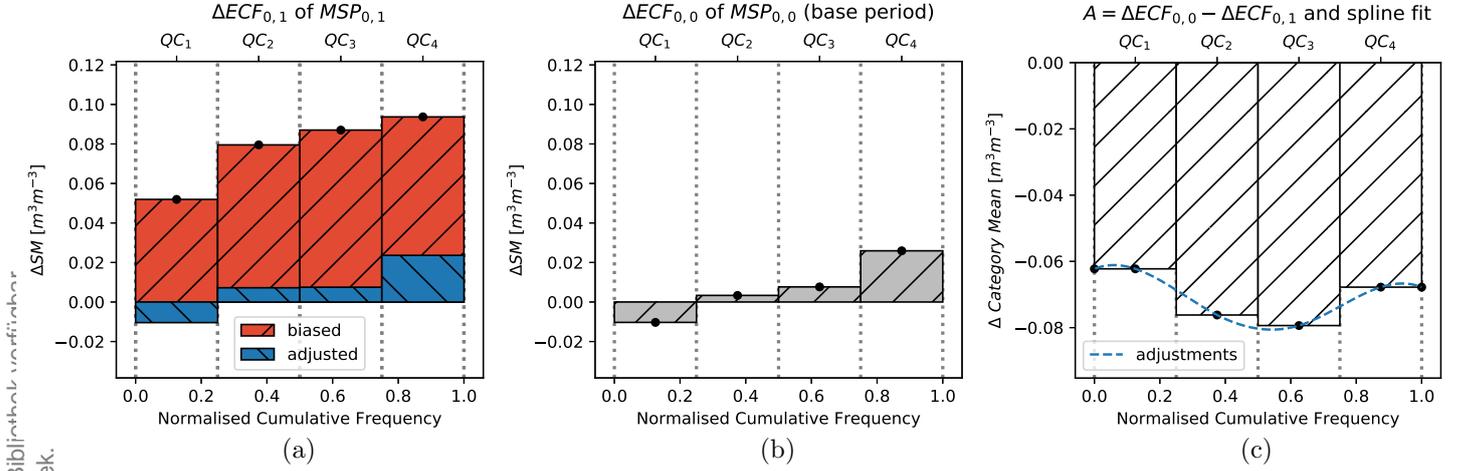


Figure 8: QCM adjustment - ΔECF for the MSP before - (a), red bars - and after (b) the transition date of the introduced break. Differences between ΔECF s (c) are used to find adjustments for all biased observations (spline interpolation, dashed blue line), and result in matching the ECF of the candidate in $MSP_{0,1}$ to the reference relative to differences in $MSP_{0,0}$ after correction - (a), blue bars.

We then calculate the category differences between the ΔECF of the two MSPs as described in (18).

$$A\left(\frac{QC_i - 0.5}{QC}\right) = \Delta ECF_{MSP_{i,0}}\left(\frac{QC_i - 0.5}{QC}\right) - \Delta ECF_{MSP_{i,1}}\left(\frac{QC_i - 0.5}{QC}\right) \quad (18)$$

where A are the corrections to fit the median CF of the candidate of the period before the break relatively to changes in the reference. To ensure a smooth transition across the categories, we fit a cubic spline to the median CFs, with $A(0)$ and $A(1)$ being equal to the first respectively last category difference value as shown in Fig. 8c. Notably, when the total number of categories $QC = 1$, all values before and after the break represent one group and adjustment results in a (constant) relative mean matching. To apply the adjustments to the candidate observations, we first find the ECF of all values to adjust (ASP_i) and apply adjustments to them, that were calculated as described in (18). Therefore we shift the extended period according to corrections from the MSPs.

The blue bars in Fig. 8a show the differences in the quantile categories between the adjusted candidate and the reference in the MSP before the break. As adjustments were found based on the extended period ASP_0 , $\Delta ECF_{0,1}$ after correction does not exactly match $\Delta ECF_{0,0}$ (in 8b), but is almost equal and differences are much smaller than before adjustment.

4.3 Evaluation

We evaluate and compare the fidelity of the described homogenisation approaches by using quantitative error metrics. In-depth validation of ESA CCI SM has been the subject of numerous studies [73], [74]. Our goal is to assess the impact of the three adjustment methods on the original data set. Validation is performed with respect to the ERA5 reanalysis as it can be considered mostly independent from the satellite SM observations and the MERRA2 reference product used in the adjustment process. We compute: Bias (difference in means), root-mean-square-difference (RMSD), unbiased root-mean-square-difference (ubRMSD), Pearson’s and Spearman’s correlation (R_P and R_S) as well as the mean-square error (MSE) and the residual sum of squares (RSS). Before calculation of the ubRMSD, ESA CCI SM datasets were scaled to ERA5 by matching their mean and standard deviation [55]. Biases between ESA CCI SM and ERA5 are expected to be dominated by discrepancies between GLDAS Noah (which is used in the production of ESA CCI SM) and ERA5. Error metrics are calculated for the single SSPs of ESA CCI SM v04.5 COMBINED from Fig. 2 as well as for the full period. Intercomparison is performed only for locations that were adjusted in any sub-period and with all three methods. Therefore a total of 41518 points was used. As ERA5 is available in quarter degree resolution, validation of quasi-located grid cells (a constant offset between the cell origins was accounted for) can be performed.

We investigate the impact of adjustment methods on long term trends in the ESA CCI SM data set by comparing changes over time between the unadjusted and adjusted versions and multiple reference products (ERA5, ERA-Interim/Land, MERRA2). The impact of adjustment methods on data set trends is of special interest, as ESA CCI SM aims to provide a data record for long term climate assessments. We follow the approach as described by Dorigo *et al.* [75] and Albergel *et al.* [76] and calculate Theil-Sen trend estimates [77] of seasonal averages together with a nonparametric Mann-Kendall test [78]–[80] to detect significant SM increase or decrease ($p < 0.05$) independent from how observations are distributed. Theil-Sen estimates provide linear fitting of the un-

derlying observations that are robust to outliers due to using the median of multiple slope estimates. Only values after 1991-08-05 are used because of the consistency of underlying data sources. Linear regression based scaling [55] was applied to all datasets before calculation of slopes with the unadjusted observations acting as the scaling reference. Trends were calculated over all observations starting in 1991 as well as for pairs of subsequent SSPs after 1991 as these were the temporal basis for the applied correction.

5 Results and Discussion

5.1 Global break detection and correction

Fig. 9 shows location and types of detected breaks for the sensor transitions at 2012-07-01 (Fig. 9 left column) and 2002-06-19 (Fig. 9 right column) before and after adjustment with the three described methods. The first transition date in 2012 contains the last SSP of the dataset (base period) and therefore represents the first iteration of adjustment performed on the candidate data. These results are not influenced by any previous correction iterations and should be easiest to reproduce. Therefore they best represent direct differences between the correction methodologies. As the sensor transition at 2011-10-05 was excluded from the correction process (to avoid sub-periods that are shorter than one year) potential breaks caused by changes in the sensor constellation from this date could affect the 2012 test results as they are included in $SSP_{0,1}$. The sensor transition in 2002 contains the largest number of detected breaks in the original data set (of both types) with respect to a relatively low testing coverage. Table 3 shows the detailed test statistics on the number of detected breaks and covered points for all seven considered sensor transition dates, before ("original") and after correcting detected breaks in ESA CCI SM with the respective method ("LMP", "HOM" and "QCM"). HOM uses mostly quadratic models (in 66% to 86% of concurrent cases, compare to SI Fig. 14). 1987-07-09 is excluded from the table as no testing is possible.

Figs. 9a and 9b show the initial test results of the original SSPs in the unadjusted dataset. For all tested transition dates the number of detected breaks in mean is 5-7 times larger than the number of variance breaks. The highest number of breaks in mean and most simultaneous (mean *and* variance) breaks are detected for the sensor transition in 2002-06-19 (29.13% respectively 3.19%, relative to the number of tested points). 1991-08-05 shows the highest relative coverage of variance breaks with 4.40%. The least breaks are detected for the sensor transition in 2010-01-15. As three of the four sensors used after 2010 were already part of the merged dataset in the preceding SSP, this indicates that the homogeneity of the dataset is affected more if more sensor products are replaced or complemented by others and less if the sensor constellation remains similar. An overall decrease in the absolute number of tested points ("Test Coverage") towards earlier periods of ESA CCI SM can be observed. This is mainly due to the smaller number of available sensors, the observation frequency and therefore the data density in these periods but also due to slightly decreasing correlation with MERRA2. This leads to the prerequisites for testing not being met and therefore no testing being possible for the first sensor transition in 1987-07-09. The number of tested

ESA CCI SM points for later periods increases from $\sim 20\%$ in 1991 up to more than 50% in 2010 and 2012.

Figs. 9c-9h show results of a separate test run after correction with the three respective methods (compare to the statistics in Table 3). The original SSPs without any extension (same as for the unadjusted data in 9a and 9b) were used. It can be seen that the number of detected breaks is significantly reduced by all three correction methods. Overall, the results for all three methods are very similar in terms of coverage and number and location of (un)removed breaks. The number of breaks in mean detected by WK is reduced by $\sim 70-80\%$ compared to the original data. Repeated testing after QCM shows the lowest number of detected breaks (of both types) after adjustment for all tested dates compared to the other methods. All methods perform better in terms of reducing breaks in mean compared to removing variance breaks. Variance breaks are reduced by 20-30% by the QCM method for all dates, but LMP and HOM show for three sensor transitions even a minor increase in FK-positive points (red numbers in Table 3). This can be the case as some breaks that were initially detected by both tests (mean *and* variance) got partly removed during earlier iterations of correction. This is due to the iterative correction approach, where one iteration might not only correct the break at its respective transition date (between MSPs) but also at the sensor transition date of the previous break, i.e. at the beginning of the adjusted period ASP_i . This explanation is supported by the fact that the number of concurrent mean *and* variance breaks ("Both") is reduced by all methods significantly, but again most for QCM. An increase in detected (only) variance breaks is therefore not necessarily an indicator of data degradation.

The remaining breaks that were detected after adjustment are either for points where preconditions were not met (and hence no adjustment was attempted) or where the adjusted values were discarded (as a bias was introduced or as a break was still detected after correction).

5.1 Global break detection and correction

Table 3: TEST STATISTICS BEFORE (ORIG.) AND AFTER ADJUSTMENT WITH LMP, HOM AND QCM

| | | Test** | <i>Break Detection</i> | | | |
|-------------------|-------|----------|------------------------|-------------|-------------|--------------|
| | | Coverage | Mean * | Variance * | Both * | No Break * |
| 1991-08-05 | orig. | 20.17 | 17.84 | <u>4.40</u> | 1.61 | 76.15 |
| | LMP | 20.14 | 5.30 | 3.75 | 0.94 | 90.01 |
| | HOM | 20.13 | 6.13 | 3.80 | 1.13 | 88.94 |
| | QCM | 20.16 | 4.39 | 3.65 | 0.66 | 91.30 |
| 1998-01-01 | orig. | 34.22 | 20.53 | 3.53 | 1.52 | 74.41 |
| | LMP | 34.15 | 5.10 | 3.37 | 0.92 | 90.61 |
| | HOM | 34.13 | 7.09 | 3.40 | 0.98 | 88.52 |
| | QCM | 34.18 | 3.87 | 2.81 | 0.52 | 92.80 |
| 2002-06-19 | orig. | 29.34 | <u>29.13</u> | 3.85 | <u>3.19</u> | 63.83 |
| | LMP | 29.31 | 9.54 | 4.16 | 1.45 | 84.85 |
| | HOM | 29.28 | 11.32 | 4.00 | 1.39 | 83.30 |
| | QCM | 29.33 | 8.34 | 3.13 | 0.90 | 87.63 |
| 2007-01-01 | orig. | 39.13 | 24.02 | 3.58 | 2.79 | 69.61 |
| | LMP | 39.09 | 8.81 | 3.73 | 1.76 | 85.70 |
| | HOM | 39.08 | 10.64 | 3.74 | 1.67 | 83.95 |
| | QCM | 39.11 | 7.38 | 3.00 | 1.06 | 88.55 |
| 2010-01-15 | orig. | 53.08 | 12.97 | 1.83 | 0.55 | 84.65 |
| | LMP | 53.04 | 4.71 | 1.70 | 0.43 | 93.17 |
| | HOM | 53.03 | 5.14 | 1.68 | 0.37 | 92.81 |
| | QCM | 53.07 | 3.86 | 1.26 | 0.25 | 94.63 |
| 2012-07-01 | orig. | 51.49 | 25.07 | 3.13 | 1.93 | 69.87 |
| | LMP | 51.48 | 7.25 | 3.23 | 1.15 | 88.37 |
| | HOM | 51.47 | 9.12 | 3.20 | 1.08 | 86.59 |
| | QCM | 51.48 | 5.64 | 2.52 | 0.67 | 91.17 |

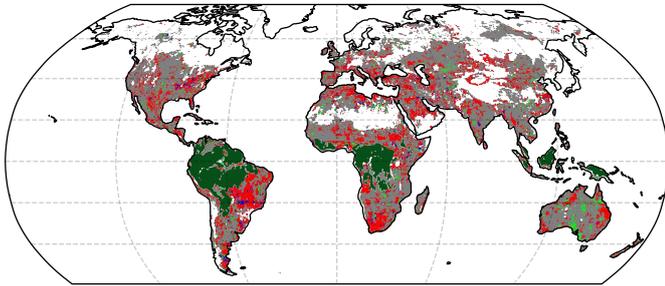
** in % of 229392 ESA CCI SM land points without dense vegetation.

* in % of tested points

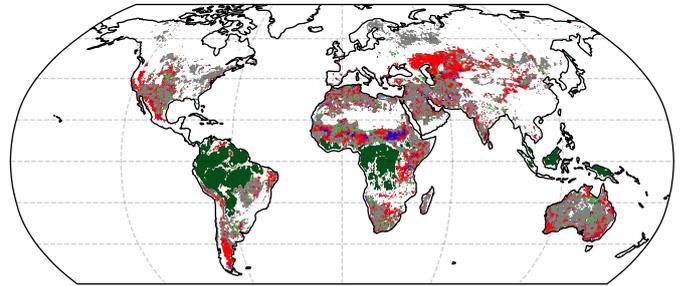
Underlined numbers indicate the highest numbers of relative breaks of each type across all transition dates, **bold** numbers indicate the best performing method in terms of reduced breaks for all transition dates. **Red** numbers indicate an increase in breaks compared to the unadjusted observations.

2012-07-01

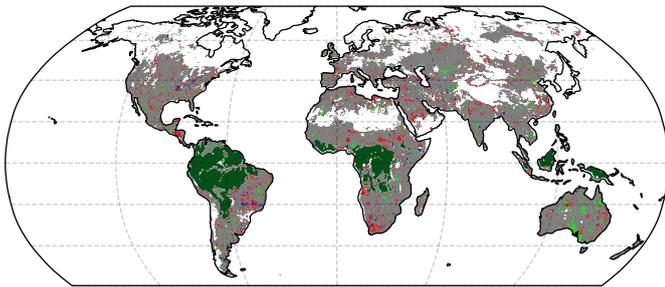
2002-06-19



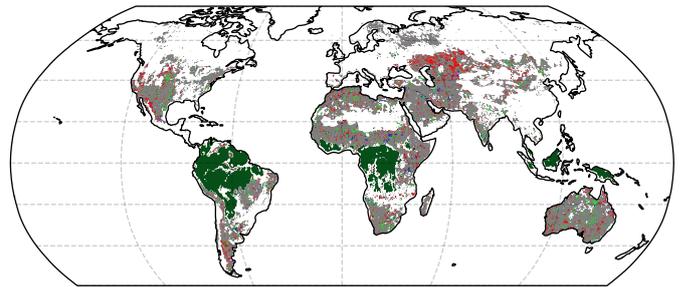
(a) 2012-07-01: Before adjustment



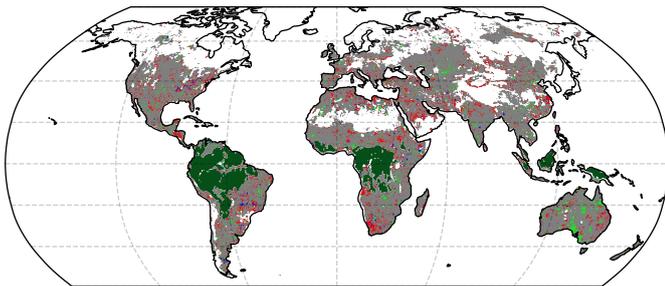
(b) 2002-06-19: Before adjustment



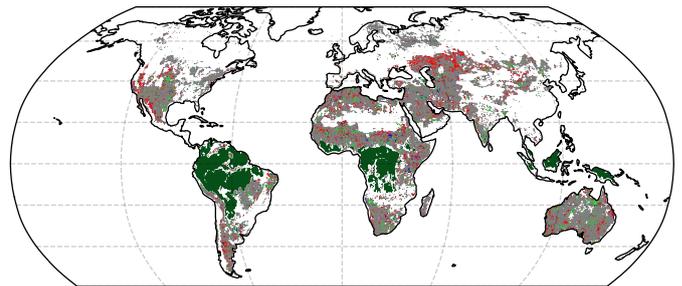
(c) 2012-07-01: After LMP adjustment



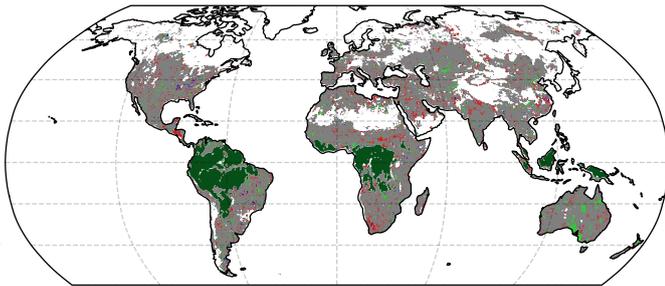
(d) 2002-06-19: After LMP adjustment



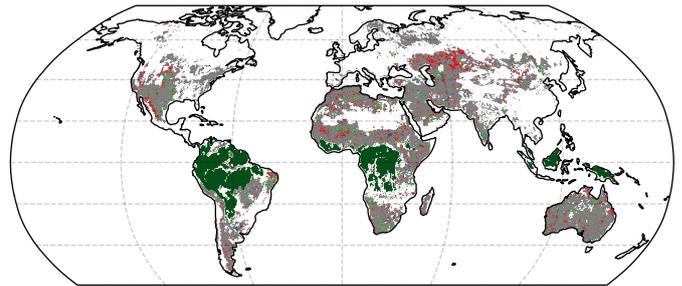
(e) 2012-07-01: After HOM adjustment



(f) 2002-06-19: After HOM adjustment



(g) 2012-07-01: After QCM adjustment



(h) 2002-06-19: After QCM adjustment

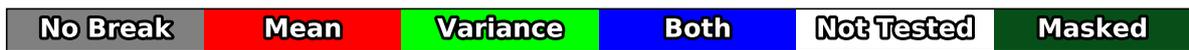


Figure 9: Break detection for ESA CCI SM v04.5 COMBINED before adjustment (a, b), after adjustment with LMP (c, d), after adjustment with HOM (e, f) and after adjustment with QCM (g, h) for two example sensor transition dates: 2012-07-01 (left column) and 2002-06-19 (right column). Plots for the remaining transition dates can be found in SI Figs. 7, 12, 17, and 21. Red areas indicate a break in mean (WK). Light Green areas indicate a shift in variance (FK). Dark blue areas define points where both tests were positive and grey areas where no (significant) break in mean *and* variance was detected. White areas indicate that testing was not performed because prerequisites are not met (see Section 4.1 and SI Fig. 8). Dark green areas are excluded according to the ESA CCI SM rainforest mask.

5.2 Analysis of intermediate results

In this section we compare some of the intermediate results of (global) break correction in more detail. Plots here only show a (representative) subset of the global results to visualise the (small) differences between the methods. The chosen region (grasslands/savannas) contains a large amount of breaks in mean and variance around the masked rainforest area. Fig. 10 shows the according subset of global break detection for the sensor transition in 1998-01-01. Fig. 10a shows results of the prerequisite checks before initial testing. Areas marked in blue were actually tested, leading to the detected homogeneities and breaks shown in Fig. 10b. Based on these results, MSPs and ASPs for adjustment were created by extending the original sensor periods. Adjustment methods are then applied, but only for points where also testing with MSPs detected a break. Comparing Figs. 10b and 10c shows that this is the case for most detected breaks. This indicates that the use of extended homogeneous sub-periods for adjustment is a viable option to increase the data size for correction without changes in the detection of a break or (less important) its type.

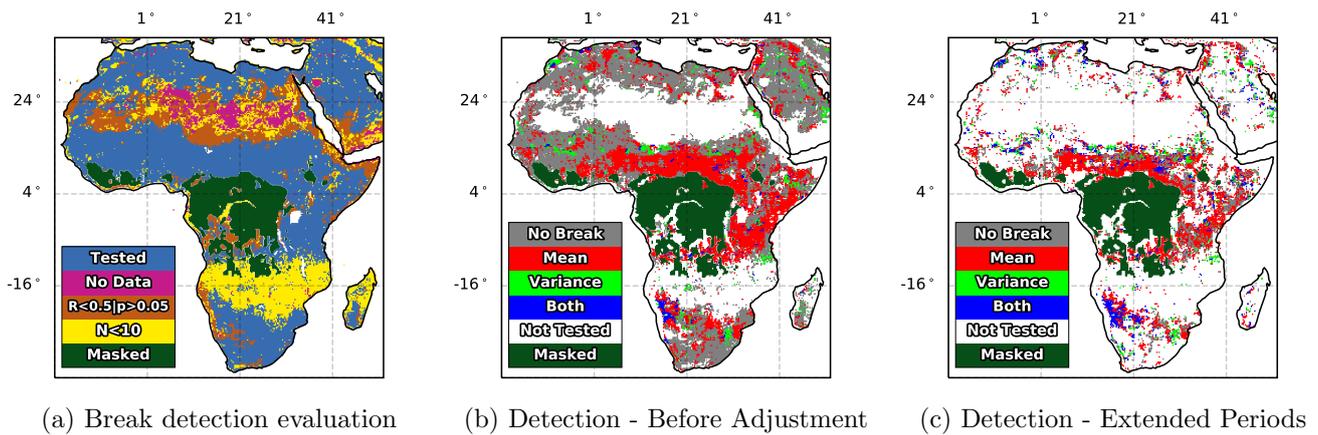


Figure 10: 1998-01-01 - subset of global results. Evaluation before initial testing (a) and detected breaks using (unextended) SSPs (b). Detected breaks after SSP extension (MSPs) (c) are the basis for break correction (see SI Figs. 9, 13 and 18 for global plots). Masked areas due to dense vegetation are excluded according to the previous figures.

Fig. 11 shows intermediate test results during correction as well as remaining breaks after correction for the same date. The left column shows detected breaks between MSPs after attempting to correct inhomogeneities as found in 10c with the three methods (LMP in line 1, HOM in line 2, and QCM in line 3). These tests represent the status

of the adjusted candidate before deciding whether to discard or accept the adjusted values. It can be seen that all three methods manage to reduce breaks in mean. For QCM a small number and for HOM a slightly larger number of breaks in mean is still detected afterwards. It can also be seen in the left column that the number of breaks in variance increases for all methods (least for QCM) compared to Fig. 10c. This shows that correction methods overall achieve to adjust the shifts in mean, but simultaneously often introduce new ones in variance instead, and hence change the type of the break instead of removing it. For LMP this indicates that the differences in the parameters of the two linear models are in fact a representation of relative mean differences, but matching observations based on slope and intercept often fails to maintain the homogeneity in variances for daily observations. Similar results can be observed for HOM (Fig. 11d). This indicates that the LOWESS fitted quantile differences between the observed and predicted candidate values include the searched correction in mean but often also affect the variance. This could be due to differences in the distributions of the candidate in the used sub-periods and/or due to how corrections are derived from interpolation of the residuals (smoothing factor α). QCM finally shows the lowest number of variance breaks in this regard. This indicates that even a small number of quantile categories ($QC \leq 4$) leads to a better performance in terms of restoring or maintaining the homogeneity in variance compared to the other two methods, albeit not in all cases.

The middle column in Fig. 11 compares adjustment output evaluations for the three methods. The three respective classes contain only points that were attempted to be adjusted (according to the left column). 1) "Accepted" contains points where correction was attempted and accepted (i.e. no break was detected anymore afterwards, output evaluations passed). The adjusted values are then stored. 2) "Break" indicates points, where after three repetitions of adjustment a break is still detected (the break type might have changed). In these cases the corrected values are rejected and discarded and the unadjusted observations within the (unsuccessfully) adjusted sub-period are restored. 3) " $> \Delta\text{Bias}$ " indicates that the difference in bias between candidate and reference in the adjusted period ASP_i and the reference period $MSP_{i,0}$ was not reduced. These corrections are then rejected as well and the original values are restored.

Finally, the right column in Fig. 11 shows test results for the adjusted products. They bring into context the coverage of adjusted and unadjusted areas for each method to a separate test run with the homogenised data set, again with the (original, unextended) SSPs. As expected the patterns of remaining breaks are similar to the evaluations in the middle column (unadjusted breaks should be detected again), but differences are expected due to the different time periods used.

5.2 Analysis of intermediate results

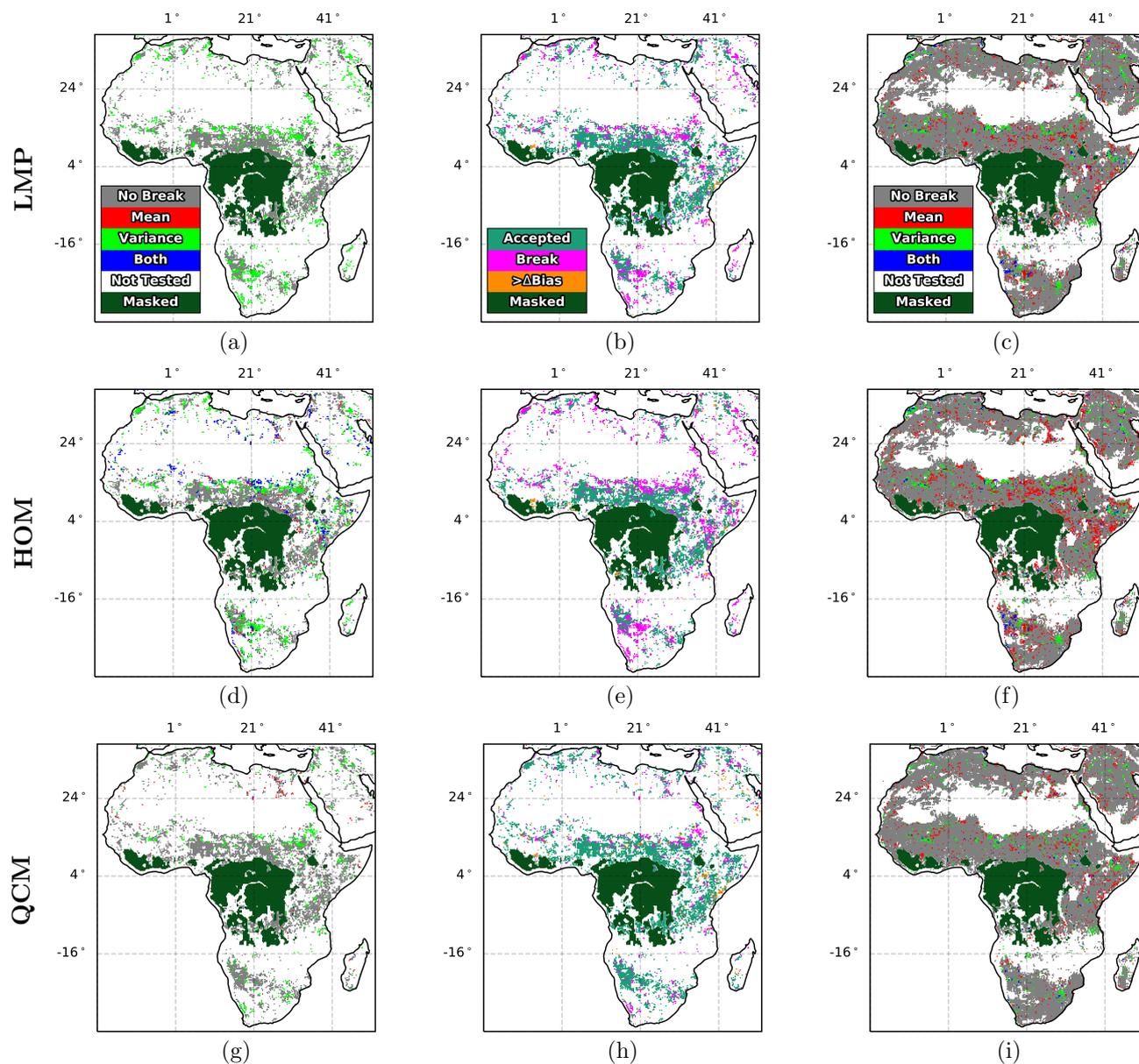


Figure 11: 1998-01-01 - Test results between MSPs after applying corrections (left column), adjustment output evaluations (middle column), and detected breaks in a separate repetition of testing between SSPs (right column). Results are shown for each of the three methods: LMP (a-c), HOM (d-f) and QCM (g-i). Densely vegetated areas are masked out as in previous plots. For the according global plots for the other transition dates, see SI Figs. 10, 11 and 15, 16 and 19, 20.

Table 4
GLOBAL STATISTICS FOR ADJUSTMENT COVERAGE

| | | Accepted* | Break* | UnCorr.* | > Δ Bias* |
|------------|-----|-----------|--------|----------|------------------|
| 1991-08-05 | LMP | 61.50 | 22.92 | 3.52 | 12.05 |
| | HOM | 59.59 | 27.48 | 1.17 | 11.76 |
| | QCM | 69.77 | 12.23 | 3.52 | 14.48 |
| 1998-01-01 | LMP | 66.79 | 26.68 | 3.22 | 3.31 |
| | HOM | 60.29 | 33.09 | 1.76 | 4.85 |
| | QCM | 78.17 | 14.74 | 3.24 | 3.86 |
| 2002-06-19 | LMP | 65.75 | 28.45 | 2.43 | 3.36 |
| | HOM | 64.33 | 31.08 | 0.91 | 3.68 |
| | QCM | 77.24 | 16.84 | 2.43 | 3.48 |
| 2007-01-01 | LMP | 59.30 | 31.30 | 2.19 | 7.21 |
| | HOM | 55.23 | 35.84 | 1.26 | 7.67 |
| | QCM | 71.77 | 17.16 | 2.18 | 8.89 |
| 2010-01-15 | LMP | 61.80 | 25.84 | 2.11 | 10.25 |
| | HOM | 61.78 | 26.39 | 0.98 | 10.85 |
| | QCM | 75.10 | 10.06 | 2.11 | 12.73 |
| 2012-07-01 | LMP | 72.50 | 20.65 | 2.20 | 4.65 |
| | HOM | 69.10 | 24.44 | 0.87 | 5.58 |
| | QCM | 82.35 | 9.60 | 2.20 | 5.85 |

* in % of detected breaks

Table 4 shows statistics for the global adjustment coverage (compare to the middle column of Fig. 11) for the three methods, i.e. the number of points where adjusted values were accepted, correction was attempted but adjusted values were discarded (" $> \Delta$ Bias" and "Break") or correction was not attempted ("UnCorr."). QCM shows the largest number of accepted adjustments, globally about 10% more than the other two methods, for which 60-70% of corrections were accepted. This corresponds with the lowest number of detected breaks after adjustment with QCM as mentioned in the discussion of Fig. 9. As expected HOM has the smallest number of points where adjustment was not attempted due to too low correlation ("UnCorr."), as for this method only the MSP after the break is evaluated, whereas for LMP and QCM R_P of both periods - $MSP_{i,0}$ and $MSP_{i,1}$ - is checked separately. For up to 15% of detected breaks adjusted values are discarded due to a relative increase in bias. This can be the case if $ASP_i \neq MSP_{i,0}$ and

could for example be due to an undetected break in the adjusted period if corrections that aim to match the MSPs are then applied to all values. The number of points for which the corrections were not accepted for this reason is similar across the three methods. Especially the transitions in 1991 and 2010 show a higher number of these cases. QCM introduces biases slightly more often than the other two methods, but this is due to more points being evaluated in this regard as less (variance) breaks remain after correction.

5.3 Impact of adjustment

5.3.1 Longest Homogeneous Period

Fig. 12 shows the length on the longest homogeneous period, i.e. the maximum period in the ESA CCI SM v04.5 COMBINED data set over which no break was detected (compare to Su *et al.* [45]). Untested sensor transitions are treated in the same way as positively tested ones (*not* homogeneous). Fig. 13 shows changes in the length of the longest homogeneous period after correction with the three respective methods. Due to the large number of removed breaks - especially in regions in Central Africa, Australia and South America - an increase in homogeneous period length can be observed. Northern regions as well as South-East Asia and Europe are less or not affected due to the overall low number of tested points there. Differences in the longest homogeneous periods between the methods can only be found for regions where the performance of adjustment methods differs, i.e. mostly in regions where variance breaks were detected. QCM (Fig. 13c) shows the longest homogeneous periods e.g. in Central Africa, North America or Australia.

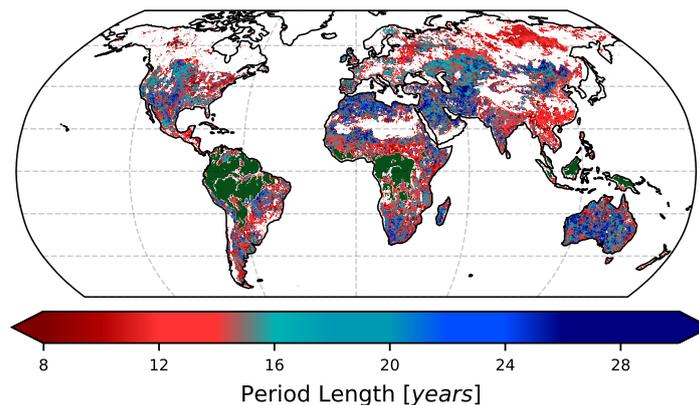


Figure 12: Length of the longest homogeneous sub-period in the unadjusted ESA CCI SM dataset (see SI Fig. 22. for according plots after correction).

5.3 Impact of adjustment

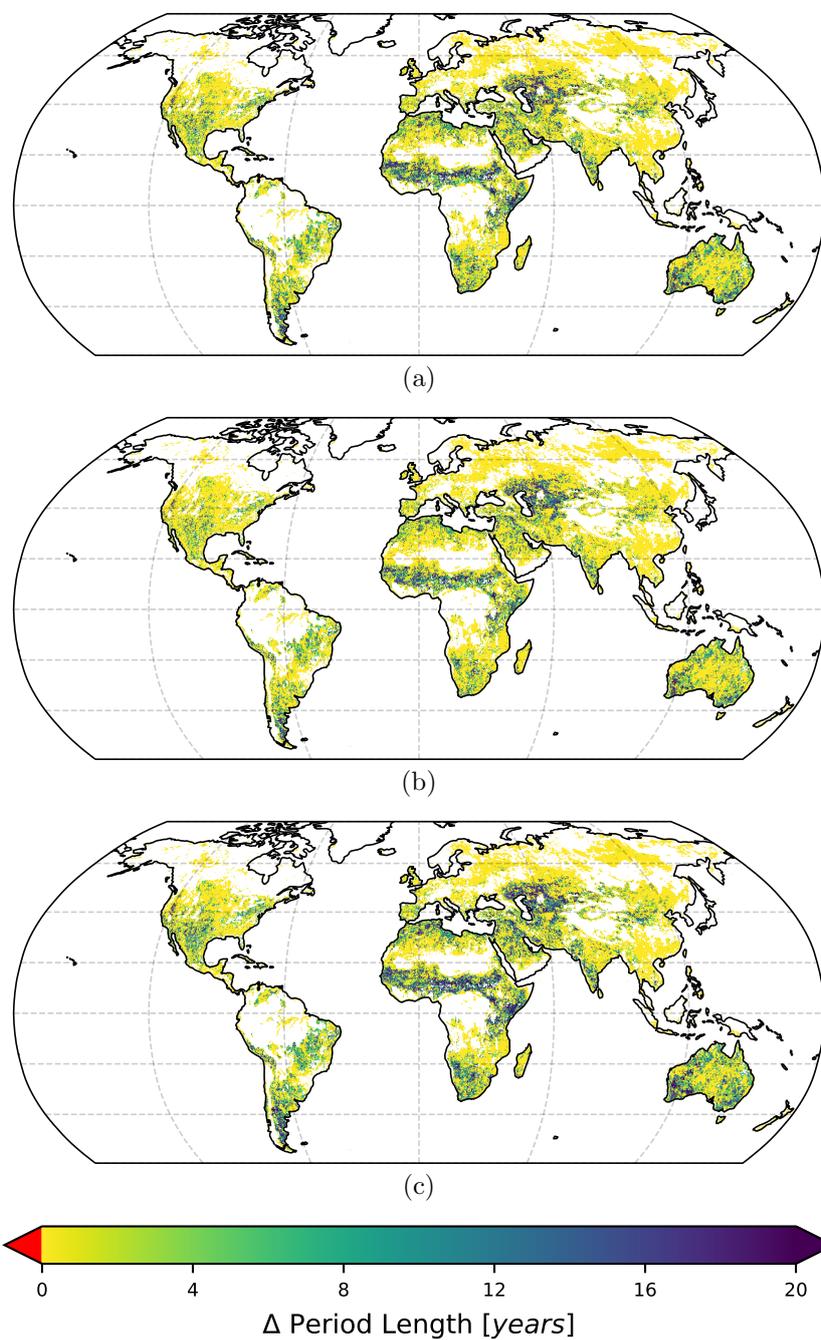


Figure 13: Increase in length of the longest homogeneous period after correction with LMP (a), HOM (b) and QCM (c).

5.3.2 Comparison against independent reanalysis data

Global comparison of validation metrics between the unadjusted and adjusted versions of ESA CCI SM v04.5 COMBINED against ERA5 SWVL1 shows only small differences between the methods. For the full period (between 1979 and 2018) we find that R_P of 41518 adjusted points increased in 55.54% of cases by up to 0.055 (for the 99th percentile) and decreased for the remaining 44.46% by down to -0.048 (for the first percentile of the differences). Slightly better results are found for QCM, where R_P for 59.95% of points increased by up to 0.057. For HOM we find a decrease in R_P for 56.11% of pixels, which indicates a slightly worse performance than the other two methods. A decrease in R_P can be observed mostly in vegetated areas - around the masked areas. As expected, we find similar results for ubRMSD. Further, QCM shows the best performance of the three methods with an overall decrease in ubRMSD with areas in Central Africa and South-America showing small increases (maps can be found in the SI Figs. 35-38). The same applies to the other metrics that were calculated, although MSE and RMSD are found to decrease in most of the before mentioned vegetated areas (e.g. in South-America). Differences in the metrics for single SSPs are found to be generally slightly more distinctive than for the full period. For example in SSP_4 ΔR_P ranges between -0.14 (HOM) and 0.08 for (LMP) and in SSP_1 $\Delta RMSD$ is between $-0.04 m^3m^{-3}$ and $0.03 m^3m^{-3}$. Compared to the other methods, QCM shows the largest improvements during SSP validation with up to 70% of pixels showing a better correspondence with ERA5 after correction. Notably we detect a slight decrease in Spearman's R (R_S) after LMP and HOM with respect to ERA5 (up to -0.3 for single points in some of the earlier SSPs) whereas changes for QCM are closer to zero or even positive. For a detailed representation of differences in the metrics for the three compared methods, we refer to SI Table I as well as the box plots and maps in SI Figs. 23-34. Overall we find that reducing the number of breaks in ESA CCI SM with the described methods has little impact on the temporal correspondence with ERA5. This indicates that the corrected relative biases within the product are small compared to the overall deviations. QCM shows the best overall performance in terms of the described metrics (shown for ubRMSD in Fig. 14).

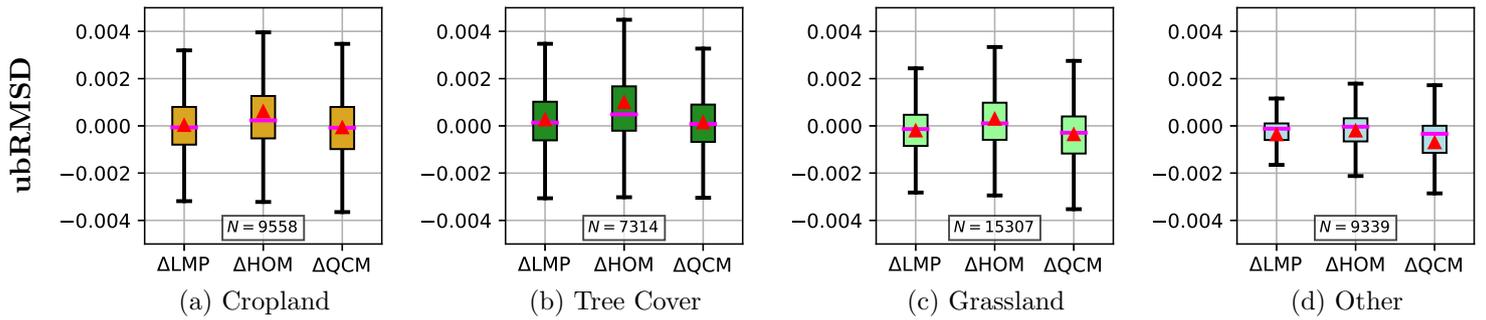


Figure 14: Difference in ubRMSD between the unadjusted and adjusted versions of ESA CCI SM v04.5 COMBINED and ERA5. Results are binned based on ESA CCI Land Cover information for the year 2010 as follows: Cropland (LC classes 10-30), Tree Cover (LC classes 40-90, 160, 170), Grassland (LC classes 120-130, 180) and all remaining points (Other). \blacktriangle indicates the mean of all values, $—$ the median, box edges represent the quartiles and whiskers $1.5 * IQR$.

5.3.3 Trend analysis

Fig. 15 shows the correspondence in (significant) SM trends in the ESA CCI SM v04.5 COMBINED data set before and after QCM adjustment. The respective actual, significant trends over the period 1991-2018 in the original and QCM adjusted ESA CCI SM data are shown in Figs. 16a and 16b (see SI Fig. 39 for the according plots for LMP and HOM, and SI Fig. 43 for the actual difference values). For comparison also trends in three reanalysis variables are shown: SWVL1 from ERA-Interim/Land in Fig. 16c (1991-2015) and from ERA5 in Fig. 16d as well as for the reference data used in the adjustment process - MERRA2 SFMC - in Fig. 16e. Differences between ESA CCI SM and reanalyses trends are evident from Fig. 16. Trends match better for the three sub-periods in SI Figs. 40-42 than for the full period.

Most significant trends are preserved during adjustment (grey areas in Fig. 15) or no significant trends are introduced (white areas in Fig. 15). There are nearly no locations in Fig. 15 where a significant drying trend changed into a significant wetting trend or vice versa. This indicates that - even though a large number of inhomogeneities was removed - the original ESA CCI SM data was influenced only little by the removed breaks and that most corrected breaks were small. This also applies to the sub-period trends. As Su *et al.* [45] pointed out, trends across inhomogeneities can be similar to those in reanalysis reference data, explaining why correction shows only limited change in this regard. Overall this strengthens the confidence in the merging scheme of ESA

CCI SM and its viability for creating a dataset for long term studies. How much single (extreme) values or noise are affected by the correction procedure is not evident from this analysis.

There are however some areas where trends that were not significant before adjustment became significant (light red and light blue areas in Fig. 15), or where significant trends are not detected anymore after adjustment (orange and turquoise areas in Fig. 15). All three adjustment methods generally slightly increase the number of detected significant trends in the data set (LMP +6%, HOM +7%, QCM +8%). Fig. 15 highlights some areas where QCM introduced significant trends. Trends from the reference data set (Fig. 16e) are adopted in most of these cases. Area (A) in Fig. 15 shows a negative trend after correction and likewise do areas (B) and (C) (compare to Figs. 16a and 16b). The latter two are (partly) aligned with the negative trends in MERRA2. In general, trends for areas around the dense vegetation mask seem to be affected more by the correction process. This could again be due to discrepancies between the MERRA2 model and ESA CCI SM in these areas (as was shown in Section 5.3.2 for the ERA5 reanalysis). The positive trends in area (D) are mostly in line with the models. In the Eastern parts of (D) more significant drying trends than in the original ESA CCI SM are detected. Areas (E), (F) and (G) show the opposite effect, as trends after correction in these areas are mostly positive. For (E) and (F) the trends are in line with MERRA2, for (G) the trends after adjustment correspond also to ERA-Interim/Land (Fig. 16c).

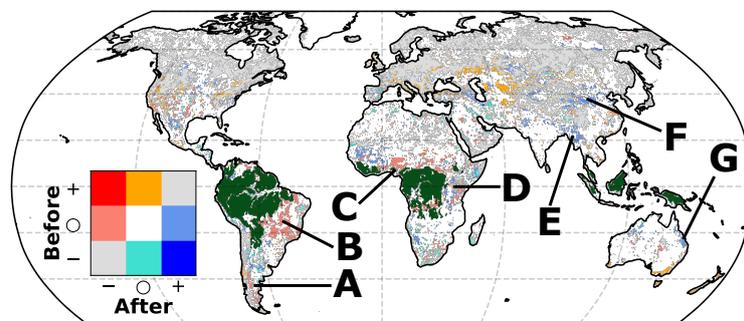


Figure 15: Sign change in significant ($p < 0.05$) SM trends after QCM adjustment. "+" indicates wetting, "-" drying and "○" non-significant trends before, respectively after adjustment. Densely vegetated areas are masked out (dark green).

5.3 Impact of adjustment

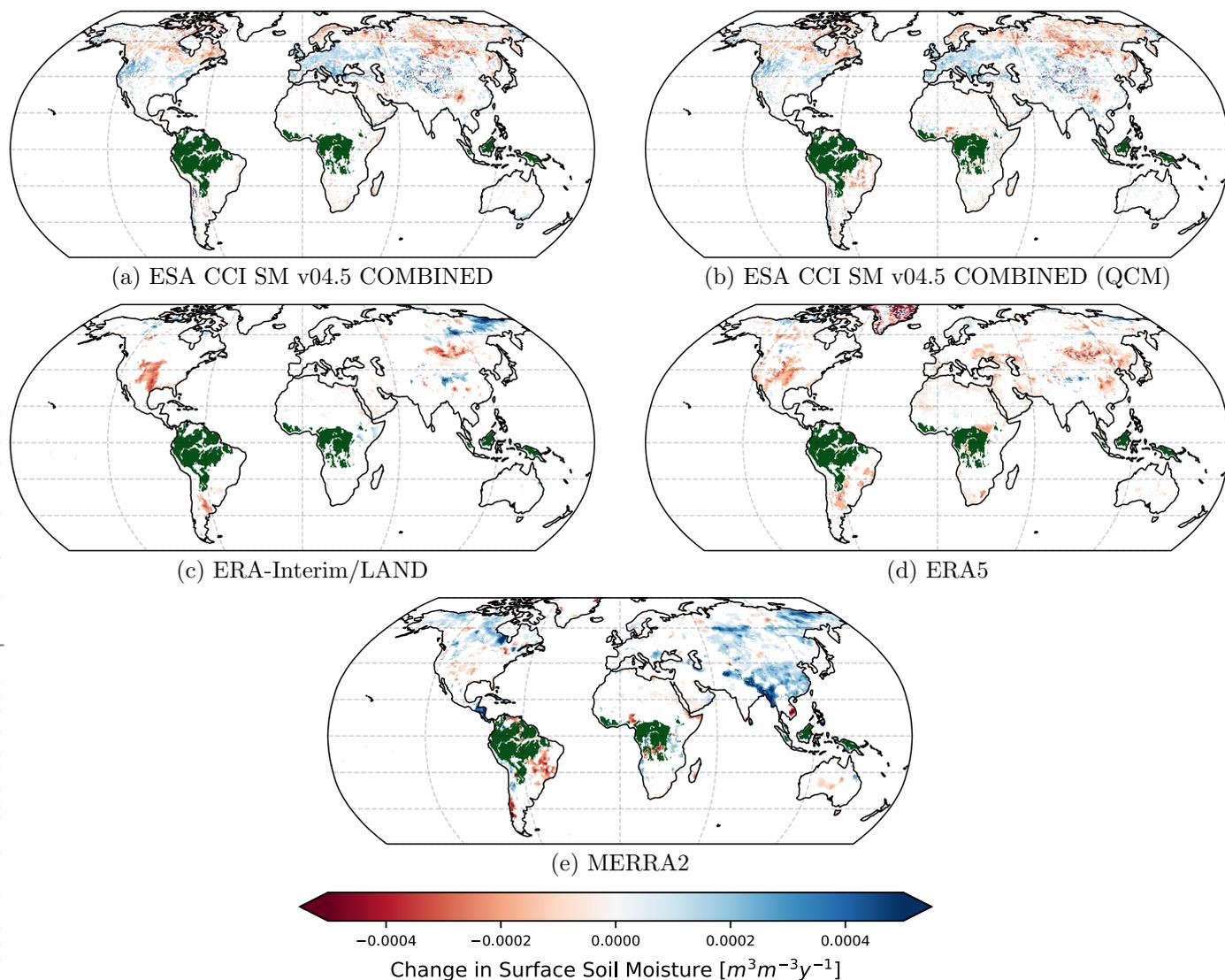


Figure 16: Significant SM trends: In the original (a) and QCM adjusted (b) ESA CCI SM v04.5 COMBINED data, in ERA-Interim/LAND SWVL1 (c), ERA5 SWVL1 (d) and MERRA2 SFMC (e).

6 Conclusion

In this study we adapted three methods to reduce structural breaks in the ESA CCI SM v04.5 COMBINED data set. Break detection was implemented in line with the preceding study by Su *et al.* [45]. Potential breaks in mean and variance as well as homogeneous and untestable transitions between sensor period subsets were detected and reduced by up to 80% with three relative (independent) approaches. All three methods depend on the use of reanalysis reference data, which was considered to be homogeneous, although it might not be [47]. Within the common framework that the methods are embedded into, extended homogeneous sub-periods from initial testing were used for quantifying detected breaks and for correction. Period extension allows applying corrections to observations between detected breaks whilst deriving them only from observations within homogeneous sub-periods. This compromise is necessary as not all sensor transitions of ESA CCI SM can be tested reliably. Repeated break detection after correction reveals that a large number of breaks (in mean) is removed by all methods. Of the three methods, Quantile Category Matching reduces the number of mean breaks the most (up to 80%) and is also the only method to reliably reduce the number of detected variance breaks. We therefore consider it the preferred method. Validation of the dataset after correction shows only little change in terms of bias, (ub)RMSD and correlation with respect to independent reanalysis reference data. In some areas break correction altered trends in ESA CCI SM. These alterations were mostly in line with the reference data set used for correction. Especially for grid cells in the Northern Hemisphere trends remained mostly unaffected. Overall we consider the limited changes in terms of error metrics and trends a confirmation of the merging approach of ESA CCI SM, as it indicates that the removed breaks were small. The impact of adjustment can be larger for spatial or temporal subsets of the data record.

Potential ways to improve the correction methods include considering different noise levels between sub-periods for quantifying breaks and during evaluation. To improve the correction of variance breaks less restrictions in terms of variability of derived corrections (e.g. with a higher number of quantile categories or a different smoothing factor for interpolation) can be explored. The framework presented here could make use of methods for joint correction [81] instead of the iterative approach that is currently implemented and make use of multiple locations (neighbourhoods) to detect and quantify breaks more reliably. The implemented iterative approach applies corrections to observations between initially detected breaks. As nonparametric tests are used to detect breaks and not all breaks can be removed, this could - to some extent - still lead to increasing the size of

detected yet not corrected breaks.

Break detection and correction should be seen as an integral part during the creation and validation of multi-sensor based climate records. This holds true for satellite-based data sets as it does for terrestrial records. Choosing suitable reference data to find breaks in satellite records is especially difficult due to the spatial and temporal coverage needed. Therefore absolute methods for break correction are of particular interest as satellite observations are known to contain features that are not represented in state of the art land surface models, such as irrigation [82].

Python code for break detection and correction from this study is publicly available on GitHub at <https://github.com/TUW-GEO/pybreaks> .

References

- [1] S. I. Seneviratne, T. Corti, E. L. Davin, M. Hirschi, E. B. Jaeger, I. Lehner, B. Orlowsky, and A. Teuling, “Investigating soil moisture–climate interactions in a changing climate: A review,” *Earth-Science Reviews*, vol. 99, no. 3, pp. 125–161, 2010. DOI: [10.1016/j.earscirev.2010.02.004](https://doi.org/10.1016/j.earscirev.2010.02.004).
- [2] H. Vereecken, J. Huisman, Y. Pachepsky, C. Montzka, J. Kruk, H. Bogena, L. Weihermüller, M. Herbst, G. Martinez, and J. Vanderborght, “On the spatio-temporal dynamics of soil moisture at the field scale,” *Journal of Hydrology*, vol. 516, pp. 76–96, Aug. 2014. DOI: [10.1016/j.jhydrol.2013.11.061](https://doi.org/10.1016/j.jhydrol.2013.11.061).
- [3] G. Bonan, *Ecological Climatology: Concepts and Applications*, 3rd ed. Cambridge University Press, 2015, ch. 1.4. DOI: [10.1017/CB09781107339200](https://doi.org/10.1017/CB09781107339200).
- [4] W. Dorigo, W. Wagner, R. Hohensinn, S. Hahn, C. Paulik, A. Xaver, A. Gruber, M. Drusch, S. Mecklenburg, P. van Oevelen, A. Robock, and T. Jackson, “The international soil moisture network: A data hosting facility for global in situ soil moisture measurements,” *Hydrology and Earth System Sciences*, vol. 15, no. 5, pp. 1675–1698, 2011. DOI: [10.5194/hess-15-1675-2011](https://doi.org/10.5194/hess-15-1675-2011).
- [5] R. Hollmann, C. J. Merchant, R. Saunders, C. Downy, M. Buchwitz, A. Cazenave, E. Chuvieco, P. Defourny, G. de Leeuw, R. Forsberg, T. Holzer-Popp, F. Paul, S. Sandven, S. Sathyendranath, M. van Roozendaal, and W. Wagner, “The esa climate change initiative: Satellite data records for essential climate variables,” *Bulletin of the American Meteorological Society*, vol. 94, no. 10, pp. 1541–1552, 2013. DOI: [10.1175/BAMS-D-11-00254.1](https://doi.org/10.1175/BAMS-D-11-00254.1).
- [6] WMO, “The global observing system for climate (gcOS): Implementation needs,” GCOS, ERA Report Series 200, 2016.
- [7] W. Wagner, K. Scipal, C. Pathe, D. Gerten, W. Lucht, and B. Rudolf, “Evaluation of the agreement between the first global remotely sensed soil moisture data with model and precipitation data,” *Journal of Geophysical Research: Atmospheres*, vol. 108, no. D19, 2003. DOI: [10.1029/2003JD003663](https://doi.org/10.1029/2003JD003663).
- [8] L. Ciabatta, C. Massari, L. Brocca, A. Gruber, C. Reimer, S. Hahn, C. Paulik, W. Dorigo, R. Kidd, and W. Wagner, “Sm2rain-cci: A new global long-term rainfall data set derived from esa cci soil moisture,” *Earth System Science Data*, vol. 10, no. 1, pp. 267–280, 2018. DOI: [10.5194/essd-10-267-2018](https://doi.org/10.5194/essd-10-267-2018).
- [9] C. Paulik, W. Dorigo, W. Wagner, and R. Kidd, “Validation of the ascat soil water index using in situ data from the international soil moisture network,” *International Journal of Applied Earth Observation and Geoinformation*, vol. 30, pp. 1–8, 2014. DOI: [10.1016/j.jag.2014.01.007](https://doi.org/10.1016/j.jag.2014.01.007).
- [10] R. M. Parinussa, T. R. H. Holmes, N. Wanders, W. Dorigo, and R. A. M. de Jeu, “A preliminary study toward consistent soil moisture from amsr2,” *Journal of Hydrometeorology*, vol. 16, no. 2, pp. 932–947, 2015. DOI: [10.1175/JHM-D-13-0200.1](https://doi.org/10.1175/JHM-D-13-0200.1).
- [11] J. M. Colston, T. Ahmed, C. Mahopo, G. Kang, M. Kosek, F. de Sousa Junior, P. S. Shrestha, E. Svensen, A. Turab, and B. Zaitchik, “Evaluating meteorological data from weather stations, and from satellites and global models for a multi-site epidemiological study,” *Environmental Research*, vol. 165, pp. 91–109, 2018. DOI: <https://doi.org/10.1016/j.envres.2018.02.027>.

- [12] H. Riebeek and R. Simmon. (2009). Catalog of earth satellite orbits, [Online]. Available: <https://earthobservatory.nasa.gov/features/OrbitsCatalog> (visited on 10/31/2019).
- [13] E. C. and J. van Zyl, "Introduction," in *Introduction to the Physics and Techniques of Remote Sensing*. John Wiley & Sons, Ltd, 2006, ch. 1, pp. 1–21. DOI: [10.1002/0471783390.ch1](https://doi.org/10.1002/0471783390.ch1).
- [14] G. P. Petropoulos, G. Ireland, and B. Barrett, "Surface soil moisture retrievals from remote sensing: Current status, products & future trends," *Physics and Chemistry of the Earth, Parts A/B/C*, vol. 83-84, pp. 36–56, 2015, Emerging science and applications with microwave remote sensing data. DOI: <https://doi.org/10.1016/j.pce.2015.02.009>.
- [15] "Ieee standard letter designations for radar-frequency bands," *IEEE P521/D1, March 2019*, pp. 1–14, Jul. 2019.
- [16] E. Schanda, *Physical Fundamentals of Remote Sensing*. Springer, 1987.
- [17] S. Misra and P. de Matthaeis, "Passive remote sensing and radio frequency interference (rfi): An overview of spectrum allocations and rfi management algorithms [technical committees]," *IEEE Geoscience and Remote Sensing Magazine*, vol. 2, no. 2, pp. 68–73, Jun. 2014. DOI: [10.1109/MGRS.2014.2320879](https://doi.org/10.1109/MGRS.2014.2320879).
- [18] A. Gruber, T. Scanlon, R. van der Schalie, W. Wagner, and W. Dorigo, "Evolution of the esacci soil moisture climate data records and their underlying merging methodology," *Earth System Science Data Discussions*, vol. 2019, pp. 1–37, 2019. DOI: [10.5194/essd-2019-21](https://doi.org/10.5194/essd-2019-21).
- [19] M. I. Skolnik, "36 - radar," in *Reference Data for Engineers (Ninth Edition)*, W. M. Middleton and M. E. V. Valkenburg, Eds., Ninth Edition, Woburn: Newnes, 2002, pp. 36-1 - 36-22. DOI: <https://doi.org/10.1016/B978-075067291-7/50038-8>.
- [20] B. Bauer-Marschallinger, C. Paulik, S. Hochstöger, T. Mistelbauer, S. Modanesi, L. Ciabatta, C. Massari, L. Brocca, and W. Wagner, "Soil moisture from fusion of scatterometer and sar: Closing the scale gap with temporal filtering," *Remote Sensing*, vol. 10, no. 7, p. 1030, Jun. 2018. DOI: [10.3390/rs10071030](https://doi.org/10.3390/rs10071030).
- [21] J. Chimot. (). Inverse model approach and challenges, [Online]. Available: <https://julien-chimot-research.blog/inverse-model-approach-and-challenges/> (visited on 10/31/2019).
- [22] p. Janssen and P. Heuberger, "Calibration of process-oriented models," *Ecological Modelling*, vol. 83, pp. 55–66, Dec. 1995. DOI: [10.1016/0304-3800\(95\)00084-9](https://doi.org/10.1016/0304-3800(95)00084-9).
- [23] M. Owe, R. de Jeu, and T. Holmes, "Multisensor historical climatology of satellite-derived global land surface moisture," *Journal of Geophysical Research: Earth Surface*, vol. 113, no. F1, 2008. DOI: [10.1029/2007JF000769](https://doi.org/10.1029/2007JF000769).
- [24] M. Owe, R. de Jeu, and J. Walker, "A methodology for surface soil moisture and vegetation optical depth retrieval using the microwave polarization difference index," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 8, pp. 1643–1654, Aug. 2001. DOI: [10.1109/36.942542](https://doi.org/10.1109/36.942542).

- [25] W. Wagner, S. Hahn, R. Kidd, T. Melzer, Z. Bartalis, S. Hasenauer, J. Figa-Saldaña, P. de Rosnay, A. Jann, S. Schneider, J. Komma, G. Kubu, K. Brugger, C. Aubrecht, J. Züger, U. Gangkofner, S. Kienberger, L. Brocca, Y. Wang, G. Blöschl, J. Eitzinger, and K. Steinnocher, “The ascats soil moisture product: A review of its specifications, validation results, and emerging applications,” *Meteorologische Zeitschrift*, vol. 22, no. 1, pp. 5–33, Feb. 2013. DOI: [10.1127/0941-2948/2013/0399](https://doi.org/10.1127/0941-2948/2013/0399).
- [26] S. Hahn and T. Melzer, “Algorithm theoretical baseline document (atbd) for metop ascats soil moisture cdr and offline products,” Eumetsat HSAF, 2017.
- [27] W. Wagner, G. Lemoine, and H. Rott, “A method for estimating soil moisture from ers scatterometer and soil data,” *Remote Sensing of Environment*, vol. 70, no. 2, pp. 191–207, 1999. DOI: [https://doi.org/10.1016/S0034-4257\(99\)00036-X](https://doi.org/10.1016/S0034-4257(99)00036-X).
- [28] W. Preimesberger, T. Scanlon, C.-H. Su, A. Gruber, and W. Dorigo, “Homogenisation of structural breaks in the global esa cci soil moisture multi-satellite climate data record,” Submitted to IEEE Transactions on Geoscience and Remote Sensing, 2019.
- [29] W. Dorigo, W. Wagner, C. Albergel, F. Albrecht, G. Balsamo, L. Brocca, D. Chung, M. Ertl, M. Forkel, A. Gruber, E. Haas, P. D. Hamer, M. Hirschi, J. Ikonen, R. de Jeu, R. Kidd, W. Lahoz, Y. Y. Liu, D. Miralles, T. Mistelbauer, N. Nicolai-Shaw, R. Parinussa, C. Pratola, C. Reimer, R. van der Schalie, S. I. Seneviratne, T. Smolander, and P. Lecomte, “Esa cci soil moisture for improved earth system understanding: State-of-the art and future directions,” *Remote Sensing of Environment*, vol. 203, pp. 185–215, 2017, Earth Observation of Essential Climate Variables. DOI: [10.1016/j.rse.2017.07.001](https://doi.org/10.1016/j.rse.2017.07.001).
- [30] Y. Liu, W. Dorigo, R. Parinussa, R. de Jeu, W. Wagner, M. McCabe, J. Evans, and A. van Dijk, “Trend-preserving blending of passive and active microwave soil moisture retrievals,” *Remote Sensing of Environment*, vol. 123, pp. 280–297, 2012. DOI: [10.1016/j.rse.2012.03.014](https://doi.org/10.1016/j.rse.2012.03.014).
- [31] A. Gruber, W. Dorigo, W. Crow, and W. Wagner, “Triple collocation-based merging of satellite soil moisture retrievals,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 12, pp. 6780–6792, Dec. 2017. DOI: [10.1109/TGRS.2017.2734070](https://doi.org/10.1109/TGRS.2017.2734070).
- [32] T. C. Peterson, D. R. Easterling, T. R. Karl, P. Groisman, N. Nicholls, N. Plummer, S. Torok, I. Auer, R. Boehm, D. Gullett, L. Vincent, R. Heino, H. Tuomenvirta, O. Mestre, T. Szentimrey, J. Salinger, E. J. Førland, I. Hanssen-Bauer, H. Alexandersson, P. Jones, and D. Parker, “Homogeneity adjustments of in situ atmospheric climate data: A review,” *International Journal of Climatology*, vol. 18, no. 13, pp. 1493–1517, 1998. DOI: [10.1002/\(SICI\)1097-0088\(19981115\)18:13<1493::AID-JOC329>3.0.CO;2-T](https://doi.org/10.1002/(SICI)1097-0088(19981115)18:13<1493::AID-JOC329>3.0.CO;2-T).
- [33] V. Conrad and L. W. Pollak, “Methods in climatology,” *Quarterly Journal of the Royal Meteorological Society*, vol. 77, no. 332, pp. 328–330, 1951. DOI: [10.1002/qj.49707733228](https://doi.org/10.1002/qj.49707733228).
- [34] B. C. Trewin and A. C. F. Trevitt, “The development of composite temperature records,” *International Journal of Climatology*, vol. 16, no. 11, pp. 1227–1242, 1996. DOI: [10.1002/\(SICI\)1097-0088\(199611\)16:11<1227::AID-JOC82>3.0.CO;2-P](https://doi.org/10.1002/(SICI)1097-0088(199611)16:11<1227::AID-JOC82>3.0.CO;2-P).
- [35] X. L. Wang, Q. H. Wen, and Y. Wu, “Penalized maximal t test for detecting undocumented mean change in climate data series,” *Journal of Applied Meteorology and Climatology*, vol. 46, no. 6, pp. 916–931, 2007. DOI: [10.1175/JAM2504.1](https://doi.org/10.1175/JAM2504.1).

- [36] J. R. Lanzante, “Resistant, robust and non-parametric techniques for the analysis of climate data: Theory and examples, including applications to historical radiosonde station data,” *International Journal of Climatology*, vol. 16, no. 11, pp. 1197–1226, 1996. DOI: [10.1002/\(SICI\)1097-0088\(199611\)16:11<1197::AID-JOC89>3.0.CO;2-L](https://doi.org/10.1002/(SICI)1097-0088(199611)16:11<1197::AID-JOC89>3.0.CO;2-L).
- [37] J. E. Pinzon and C. J. Tucker, “A non-stationary 1981–2012 avhrr ndvi3g time series,” *Remote Sensing*, vol. 6, no. 8, pp. 6929–6960, 2014. DOI: [10.3390/rs6086929](https://doi.org/10.3390/rs6086929).
- [38] S. Brinckmann, J. Trentmann, and B. Ahrens, “Homogeneity analysis of the cm saf surface solar irradiance dataset derived from geostationary satellite observations,” *Remote Sensing*, vol. 6, no. 1, pp. 352–378, 2014. DOI: [10.3390/rs6010352](https://doi.org/10.3390/rs6010352).
- [39] M. Schroeder, M. Lockhoff, J. M. Forsythe, H. Q. Cronk, T. H. Vonder Haar, and R. Bennartz, “The gewex water vapor assessment: Results from intercomparison, trend, and homogeneity analysis of total column water vapor,” *Journal of Applied Meteorology and Climatology*, vol. 55, no. 7, pp. 1633–1649, 2016. DOI: [10.1175/JAMC-D-15-0304.1](https://doi.org/10.1175/JAMC-D-15-0304.1).
- [40] M. Hulme, “Estimating global changes in precipitation,” *Weather*, vol. 50, no. 2, pp. 34–42, 1995. DOI: [10.1002/j.1477-8696.1995.tb06074.x](https://doi.org/10.1002/j.1477-8696.1995.tb06074.x).
- [41] H. Tuomenvirta, “Homogeneity adjustments of temperature and precipitation series—finnish and nordic data,” *International Journal of Climatology*, vol. 21, no. 4, pp. 495–506, 2001. DOI: [10.1002/joc.616](https://doi.org/10.1002/joc.616).
- [42] L. A. Vincent, X. Zhang, B. R. Bonsal, and W. D. Hogg, “Homogenization of daily temperatures over canada,” *Journal of Climate*, vol. 15, no. 11, pp. 1322–1334, 2002. DOI: [10.1175/1520-0442\(2002\)015<1322:HODT0C>2.0.CO;2](https://doi.org/10.1175/1520-0442(2002)015<1322:HODT0C>2.0.CO;2).
- [43] M. Rodell, P. R. Houser, U. Jambor, J. Gottschalck, K. Mitchell, C.-J. Meng, K. Arsenault, B. Cosgrove, J. Radakovich, M. Bosilovich, J. K. Entin, J. P. Walker, D. Lohmann, and D. Toll, “The global land data assimilation system,” *Bulletin of the American Meteorological Society*, vol. 85, no. 3, pp. 381–394, 2004. DOI: [10.1175/BAMS-85-3-381](https://doi.org/10.1175/BAMS-85-3-381).
- [44] T. J. Jackson, T. J. Schmugge, and J. R. Wang, “Passive microwave sensing of soil moisture under vegetation canopies,” *Water Resources Research*, vol. 18, no. 4, pp. 1137–1142, 1982. DOI: [10.1029/WR018i004p01137](https://doi.org/10.1029/WR018i004p01137).
- [45] C.-H. Su, D. Ryu, W. Dorigo, S. Zwieback, A. Gruber, C. Albergel, R. H. Reichle, and W. Wagner, “Homogeneity of a global multisatellite soil moisture climate data record,” *Geophysical Research Letters*, vol. 43, no. 21, pp. 11, 245–11, 252, 2016. DOI: [10.1002/2016GL070458](https://doi.org/10.1002/2016GL070458).
- [46] C. T. Peterson and D. Easterling, “Creation of homogeneous composite climatological reference series,” *International Journal of Climatology*, vol. 14, pp. 671–679, Jul. 1994. DOI: [10.1002/joc.3370140606](https://doi.org/10.1002/joc.3370140606).
- [47] A. Sterl, “On the (in)homogeneity of reanalysis products,” *Journal of Climate*, vol. 17, no. 19, pp. 3866–3873, 2004. DOI: [10.1175/1520-0442\(2004\)017<3866:OTI0RP>2.0.CO;2](https://doi.org/10.1175/1520-0442(2004)017<3866:OTI0RP>2.0.CO;2).
- [48] C. Azorin-Molina, S. M. Vicente-Serrano, T. R. McVicar, S. Jerez, A. Sanchez-Lorenzo, J.-I. López-Moreno, J. Revuelto, R. M. Trigo, J. A. Lopez-Bustins, and F. Espírito-Santo, “Homogenization and assessment of observed near-surface wind speed trends over spain and portugal, 1961–2011,” *Journal of Climate*, vol. 27, no. 10, pp. 3692–3712, 2014. DOI: [10.1175/JCLI-D-13-00652.1](https://doi.org/10.1175/JCLI-D-13-00652.1).

- [49] K. W. Potter, "Illustration of a new test for detecting a shift in mean in precipitation series," *Monthly Weather Review*, vol. 109, no. 9, pp. 2040–2045, 1981. DOI: [10.1175/1520-0493\(1981\)109<2040:IOANTF>2.0.CO;2](https://doi.org/10.1175/1520-0493(1981)109<2040:IOANTF>2.0.CO;2).
- [50] R. Gelaro, W. McCarty, M. J. Suárez, R. Todling, A. Molod, L. Takacs, C. A. Randles, A. Darmenov, M. G. Bosilovich, R. Reichle, *et al.*, "The modern-era retrospective analysis for research and applications, version 2 (merra-2)," *Journal of Climate*, vol. 30, no. 14, pp. 5419–5454, 2017.
- [51] GMAO. (2015). Merra-2 tavg1_2d_lnd_nx: 2d 1-hourly time-averaged single-level assimilation, land surface diagnostics v5.12.4. English, [Online]. Available: <https://disc.gsfc.nasa.gov>.
- [52] R. H. Reichle, R. D. Koster, G. J. M. De Lannoy, B. A. Forman, Q. Liu, S. P. P. Mahanama, and A. Touré, "Assessment and enhancement of merra land surface hydrology estimates," *Journal of Climate*, vol. 24, no. 24, pp. 6322–6338, 2011. DOI: [10.1175/JCLI-D-10-05033.1](https://doi.org/10.1175/JCLI-D-10-05033.1).
- [53] R. H. Reichle, C. S. Draper, Q. Liu, M. Girotto, S. P. P. Mahanama, R. D. Koster, and G. J. M. De Lannoy, "Assessment of merra-2 land surface hydrology estimates," *Journal of Climate*, vol. 30, no. 8, pp. 2937–2960, 2017. DOI: [10.1175/JCLI-D-16-0720.1](https://doi.org/10.1175/JCLI-D-16-0720.1).
- [54] Y. Liu, R. Parinussa, W. Dorigo, R. de Jeu, W. Wagner, A. van Dijk, M. McCabe, and J. Evans, "Developing an improved soil moisture dataset by blending passive and active microwave satellite-based retrievals," *Hydrology and Earth System Sciences*, vol. 15, Nov. 2010. DOI: [10.5194/hessd-7-6699-2010](https://doi.org/10.5194/hessd-7-6699-2010).
- [55] C. Paulik, A. Plocon, S. Hahn, T. Mistelbauer, M. Schmitzer, A. Gruber, I. Teubner, and C. Reimer, *Tuw-geo/pytesmo: V0.6.10*, Apr. 2018. DOI: [10.5281/zenodo.1215760](https://doi.org/10.5281/zenodo.1215760).
- [56] C. Paulik, W. Preimesberger, M. Schmitzer, and S. Hahn, *Tuw-geo/ecmwf_models: V0.5*, Jun. 2019. DOI: [10.5281/zenodo.3245388](https://doi.org/10.5281/zenodo.3245388).
- [57] H. Hersbach, P. de Rosnay, B. Bell, D. Schepers, A. Simmons, C. Soci, S. Abdalla, M. Alonso-Balmaseda, G. Balsamo, P. Bechtold, P. Berrisford, J.-R. Bidlot, E. de Boissésou, M. Bonavita, P. Browne, R. Buizza, P. Dahlgren, D. Dee, R. Dragani, M. Diamantakis, J. Flemming, R. Forbes, A. J. Geer, T. Haiden, E. Hólm, L. Haimberger, R. Hogan, A. Horányi, M. Janiskova, P. Laloyaux, P. Lopez, J. Muñoz-Sabater, C. Peubey, R. Radu, D. Richardson, J. Thépaut, F. Vitart, X. Yang, E. Zsótér, and H. Zuo, "Operational global reanalysis: Progress, future directions and synergies with nwp," ECMWF, ERA Report Series 27, Dec. 2018. DOI: [10.21957/tkic6g3wm](https://doi.org/10.21957/tkic6g3wm).
- [58] D. P. Dee, S. M. Uppala, A. J. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M. A. Balmaseda, G. Balsamo, P. Bauer, P. Bechtold, A. C. M. Beljaars, L. van de Berg, J. Bidlot, N. Bormann, C. Delsol, R. Dragani, M. Fuentes, A. J. Geer, L. Haimberger, S. B. Healy, H. Hersbach, E. V. Hólm, L. Isaksen, P. Källberg, M. Köhler, M. Matricardi, A. P. McNally, B. M. Monge-Sanz, J.-J. Morcrette, B.-K. Park, C. Peubey, P. de Rosnay, C. Tavolato, J.-N. Thépaut, and F. Vitart, "The era-interim reanalysis: Configuration and performance of the data assimilation system," *Quarterly Journal of the Royal Meteorological Society*, vol. 137, no. 656, pp. 553–597, 2011. DOI: [10.1002/qj.828](https://doi.org/10.1002/qj.828).
- [59] G. Balsamo, C. Albergel, A. Beljaars, S. Boussetta, E. Brun, H. Cloke, D. Dee, E. Dutra, J. Muñoz-Sabater, F. Pappenberger, P. de Rosnay, T. Stockdale, and F. Vitart, "Era-interim/land: A global land surface reanalysis data set," *Hydrology and Earth System Sciences*, vol. 19, no. 1, pp. 389–407, 2015. DOI: [10.5194/hess-19-389-2015](https://doi.org/10.5194/hess-19-389-2015).

- [60] H. Alexandersson and A. Moberg, "Homogenization of swedish temperature data. part i: Homogeneity test for linear trends," *International Journal of Climatology*, vol. 17, no. 1, pp. 25–34, 1997. DOI: [10.1002/\(SICI\)1097-0088\(199701\)17:1<25::AID-JOC103>3.0.CO;2-J](https://doi.org/10.1002/(SICI)1097-0088(199701)17:1<25::AID-JOC103>3.0.CO;2-J).
- [61] H. B. Mann and D. R. Whitney, "On a test of whether one of two random variables is stochastically larger than the other," *The Annals of Mathematical Statistics*, vol. 18, no. 1, pp. 50–60, Mar. 1947. DOI: [10.1214/aoms/1177730491](https://doi.org/10.1214/aoms/1177730491).
- [62] J. W. Pratt, "Remarks on zeros and ties in the wilcoxon signed rank procedures," *Journal of the American Statistical Association*, vol. 54, no. 287, pp. 655–667, 1959. DOI: [10.1080/01621459.1959.10501526](https://doi.org/10.1080/01621459.1959.10501526).
- [63] M. Fligner and T. J. Killeen, "Distribution-free two-sample tests for scale," *Journal of the American Statistical Association*, vol. 71, no. 353, pp. 210–213, 1976. DOI: [10.1080/01621459.1976.10481517](https://doi.org/10.1080/01621459.1976.10481517).
- [64] H. Tuomenvirta and H. Alexandersson, "Adjustment of apparent changes in variability of temperature time series," *6th International Meeting on Statistical Climatology*, pp. 443–446, 1995.
- [65] J. Sheng and F. Zwiers, "An improved scheme for time-dependent boundary conditions in atmospheric general circulation models," *Climate Dynamics*, vol. 14, no. 7, pp. 609–613, Jun. 1998. DOI: [10.1007/s003820050244](https://doi.org/10.1007/s003820050244).
- [66] P. M. Della-Marta and H. Wanner, "A method of homogenizing the extremes and mean of daily temperature measurements," *Journal of Climate*, vol. 19, no. 17, pp. 4179–4197, 2006. DOI: [10.1175/JCLI3855.1](https://doi.org/10.1175/JCLI3855.1).
- [67] J. R. M. Hosking, "L-moments: Analysis and estimation of distributions using linear combinations of order statistics," *Journal of the Royal Statistical Society*, vol. 52, no. 1, pp. 105–124, 1990.
- [68] F. J. Massey Jr., "The kolmogorov-smirnov test for goodness of fit," *Journal of the American Statistical Association*, vol. 46, no. 253, pp. 68–78, 1951. DOI: [10.1080/01621459.1951.10500769](https://doi.org/10.1080/01621459.1951.10500769).
- [69] W. S. Cleveland, "Robust locally weighted regression and smoothing scatterplots," *Journal of the American Statistical Association*, vol. 74, no. 368, pp. 829–836, 1979. DOI: [10.1080/01621459.1979.10481038](https://doi.org/10.1080/01621459.1979.10481038).
- [70] S. Seabold and J. Perktold, "Statsmodels: Econometric and statistical modeling with python," in *9th Python in Science Conference*, 2010.
- [71] X. L. Wang, H. Chen, Y. Wu, Y. Feng, and Q. Pu, "New techniques for the detection and adjustment of shifts in daily precipitation data series," *Journal of Applied Meteorology and Climatology*, vol. 49, no. 12, pp. 2416–2436, 2010. DOI: [10.1175/2010JAMC2376.1](https://doi.org/10.1175/2010JAMC2376.1).
- [72] L. A. Vincent, X. L. Wang, E. J. Milewska, H. Wan, F. Yang, and V. Swail, "A second generation of homogenized canadian monthly surface air temperature for climate trend analysis," *Journal of Geophysical Research: Atmospheres*, vol. 117, no. D18, 2012. DOI: [10.1029/2012JD017859](https://doi.org/10.1029/2012JD017859).
- [73] W. Dorigo, A. Gruber, R. De Jeu, W. Wagner, T. Stacke, A. Loew, C. Albergel, L. Brocca, D. Chung, R. Parinussa, and R. Kidd, "Evaluation of the esa cci soil moisture product using ground-based observations," *Remote Sensing of Environment*, vol. 162, pp. 380–395, 2015. DOI: [10.1016/j.rse.2014.07.023](https://doi.org/10.1016/j.rse.2014.07.023).

- [74] A. McNally, S. Shukla, K. R. Arsenault, S. Wang, C. D. Peters-Lidard, and J. P. Verdin, "Evaluating esa cci soil moisture in east africa," *International Journal of Applied Earth Observation and Geoinformation*, vol. 48, pp. 96–109, 2016, Advances in the Validation and Application of Remotely Sensed Soil Moisture - Part 2. DOI: [10.1016/j.jag.2016.01.001](https://doi.org/10.1016/j.jag.2016.01.001).
- [75] W. Dorigo, R. de Jeu, D. Chung, R. Parinussa, Y. Liu, W. Wagner, and D. Fernández-Prieto, "Evaluating global trends (1988–2010) in harmonized multi-satellite surface soil moisture," *Geophysical Research Letters*, vol. 39, no. 18, 2012. DOI: [10.1029/2012GL052988](https://doi.org/10.1029/2012GL052988).
- [76] C. Albergel, W. Dorigo, R. H. Reichle, G. Balsamo, P. de Rosnay, J. Muñoz-Sabater, L. Isaksen, R. de Jeu, and W. Wagner, "Skill and global trend analysis of soil moisture from reanalyses and microwave remote sensing," *Journal of Hydrometeorology*, vol. 14, no. 4, pp. 1259–1277, 2013. DOI: [10.1175/JHM-D-12-0161.1](https://doi.org/10.1175/JHM-D-12-0161.1).
- [77] H. Theil, "A rank-invariant method of linear and polynomial regression analysis," in *Henri Theils contributions to economics and econometrics*, Springer, 1992, pp. 345–381.
- [78] H. B. Mann, "Nonparametric tests against trend," *Econometrica*, vol. 13, no. 3, pp. 245–259, 1945.
- [79] M. G. Kendall, "Rank correlation methods.," 1948.
- [80] R. O. Gilbert, *Statistical methods for environmental pollution monitoring*. John Wiley & Sons, 1987.
- [81] H. Caussinus and O. Mestre, "Detection and correction of artificial shifts in climate series," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 53, no. 3, pp. 405–425, 2004. DOI: [10.1111/j.1467-9876.2004.05155.x](https://doi.org/10.1111/j.1467-9876.2004.05155.x).
- [82] F. Zaussinger, W. Dorigo, A. Gruber, A. Tarpanelli, P. Filippucci, and L. Brocca, "Estimating irrigation water use over the contiguous united states by combining satellite and reanalysis soil moisture data," *Hydrology and Earth System Sciences*, vol. 23, no. 2, pp. 897–923, 2019. DOI: [10.5194/hess-23-897-2019](https://doi.org/10.5194/hess-23-897-2019).

Supporting Information

Homogenisation of structural breaks in the global ESA CCI Soil Moisture multi-satellite climate data record

The following time series plots show the synthetic break in North Louisiana (Lat: 32.875°, Lon: -91.625°), which is used to demonstrate the correction methods.

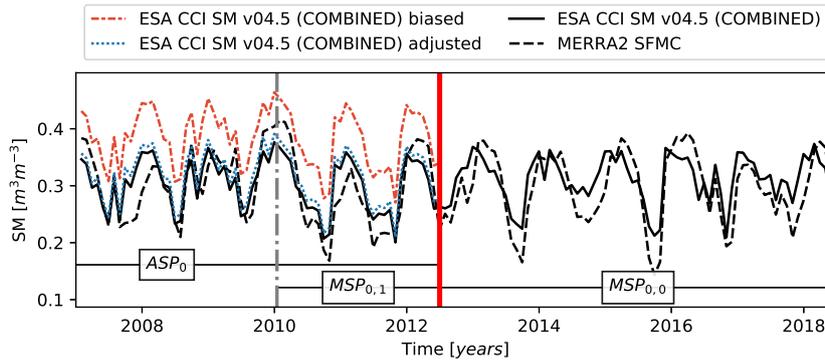


Fig. 1: *LMP adjustment*: Example for an artificially introduced break (vertical red line and dashed red time series) in a homogeneous temporal subset of an ESA CCI SM v04.5 COMBINED candidate series (solid black time series). The break is removed using LMP adjustment (dashed blue time series) using the nearest MERRA2 SFMC simulation (dashed grey time series). The original and adjusted daily values were resampled via monthly means for plotting purpose.

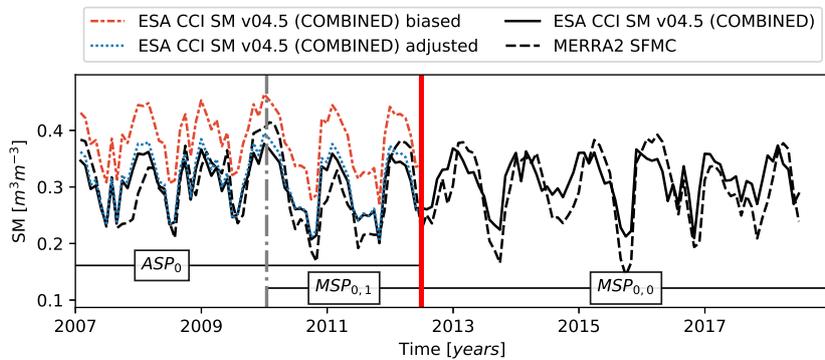


Fig. 2: *HOM adjustment*: The same artificial break as for Fig. 1. Correction is performed with HOM (dashed blue time series) to match the biased part (dashed red time series) of the candidate relatively to the reference (dashed grey time series).

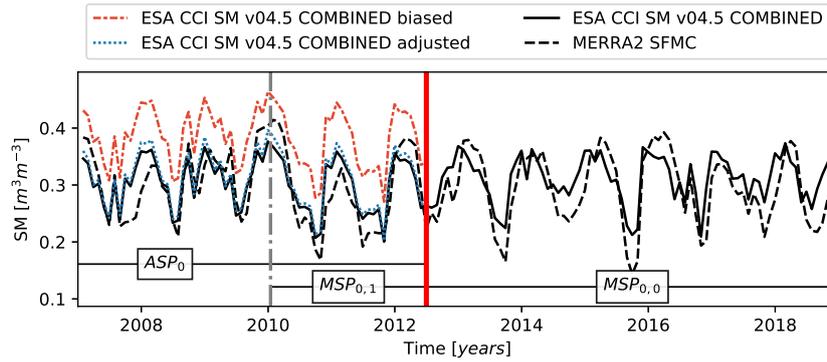


Fig. 3: *QCM adjustment*: Same artificial break as in Figs. 1 and 2. The artificially biased observations (dashed red time series) is corrected using the QCM method (dashed blue time series) with respect to relative changes in the reference (dashed grey time series).

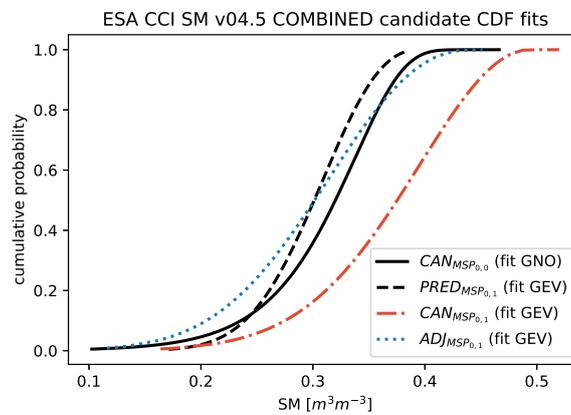


Fig. 4: *HOM adjustment*: Best fitting candidate CDFs of observed/biased ($CAN_{MSP_{0,1}}$, red line) and predicted ($PRED_{MSP_{0,1}}$, black dashed line) candidate before the date of the introduced break and of the observed candidate after the break introduction date ($CAN_{MSP_{0,0}}$, solid black line). The CDF of the adjusted candidate ($ADJ_{MSP_{0,1}}$, dotted blue line) is shown for comparison purpose only. For the example case CDFs were best matched by a Generalized Normal (GNO) respectively a Generalized Extreme Value (GEV) distribution based on the KS test.

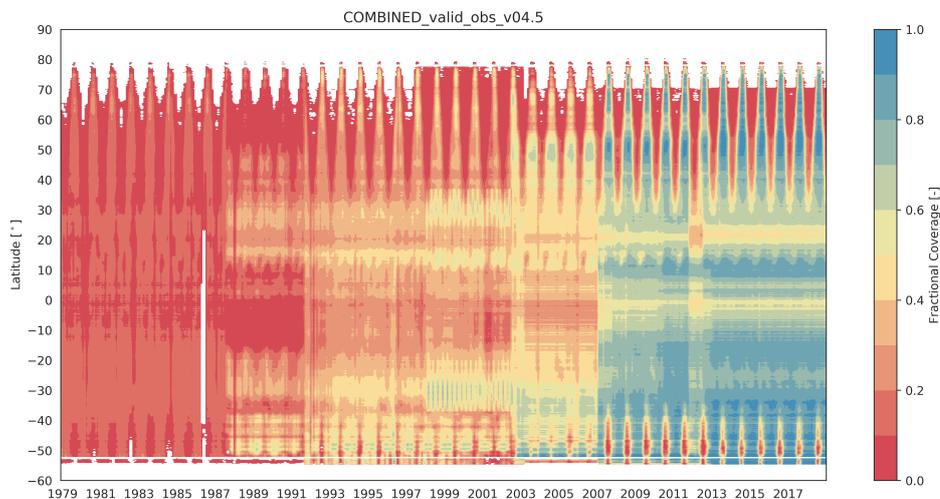


Fig. 5: Latitude-Time plot of fractional coverage of number of available observations in the ESA CCI SM v04.5 COMBINED data set.

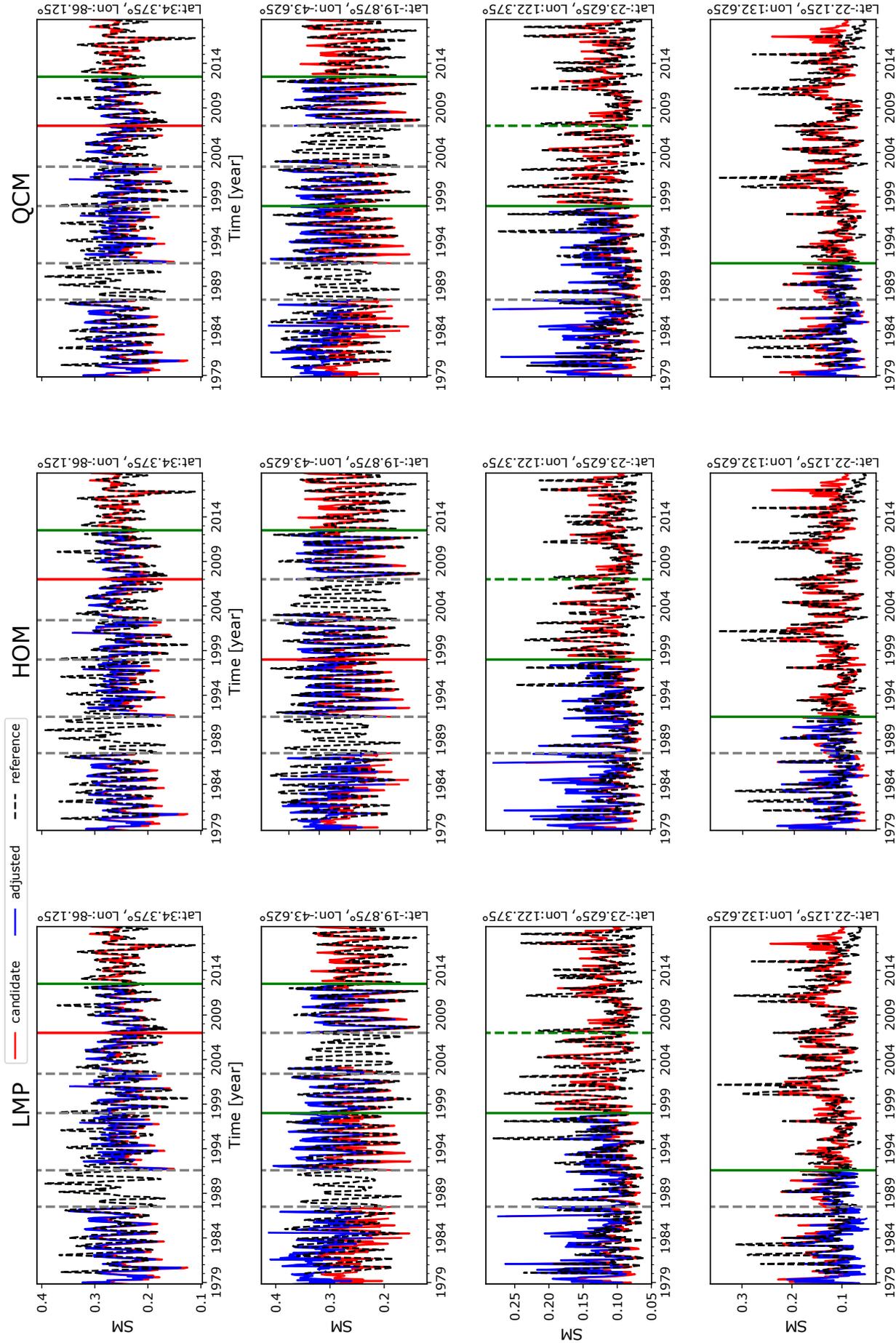


Fig. 6: Four adjustment examples for real breaks in ESA CCI SM v04.5 COMBINED with MERRA2 SFMC as the reference. Each column represent one of the three adjustment methods.

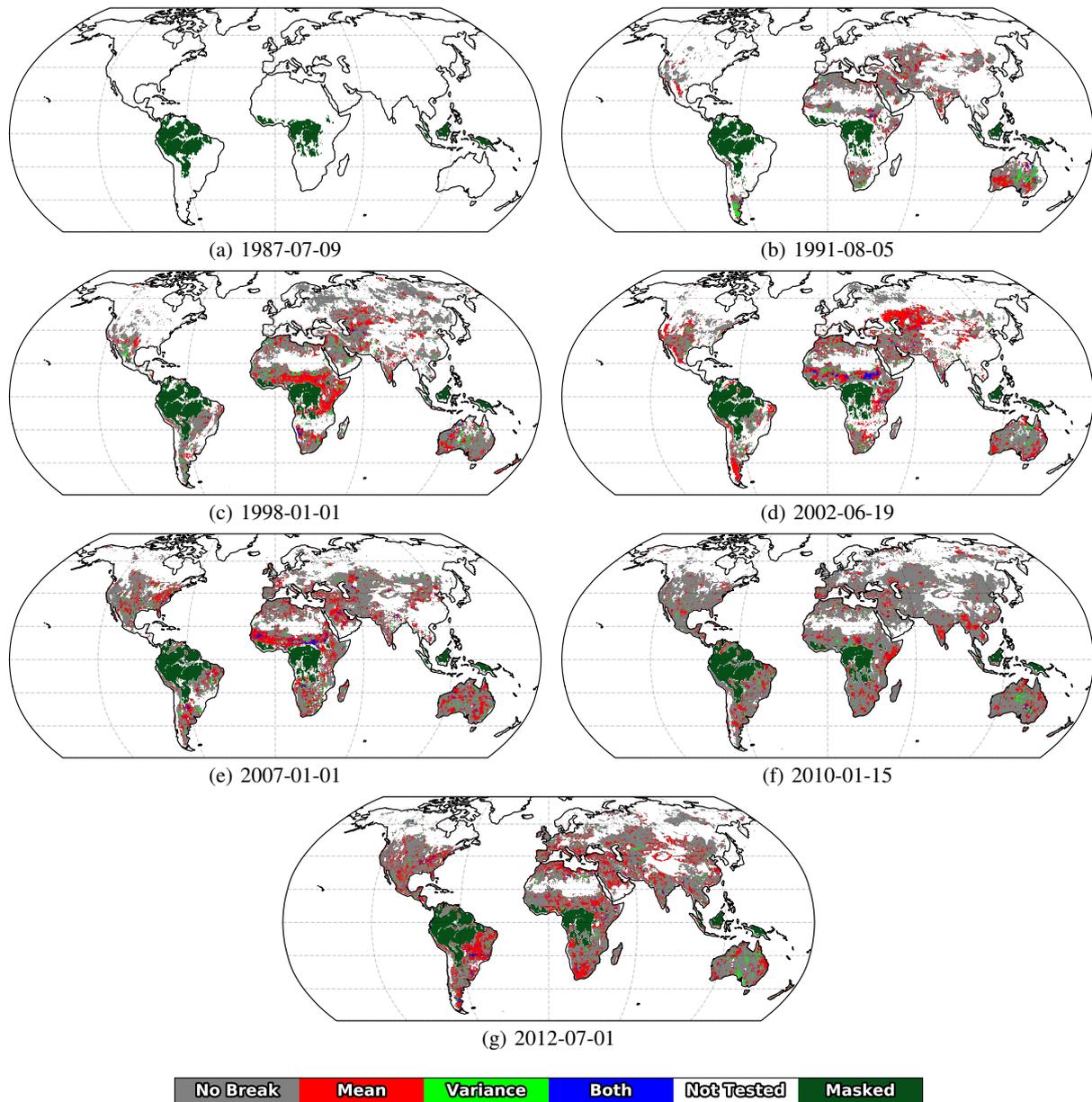


Fig. 7: Detected breaks during initial testing (with original sensor sub-periods) in the ESA CCI SM v04.5 COMBINED dataset with MERRA2 SFMC as the reference. The first sensor transition date is never tested. Dark green areas are masked due to dense vegetation.

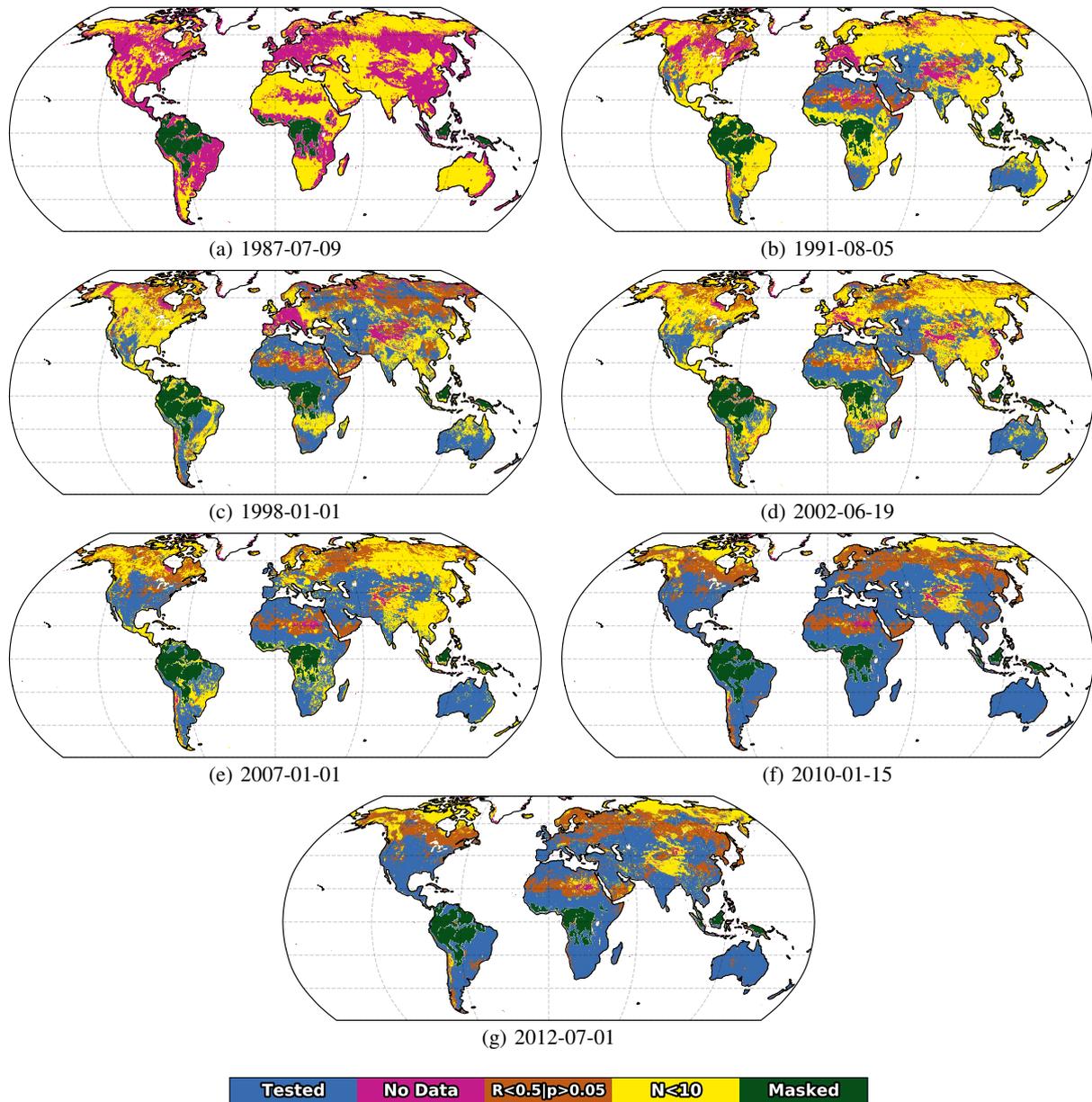


Fig. 8: Coverage of initial break testing (with original sensor periods) in the ESA CCI SM v04.5 COMBINED dataset with MERRA2 SFMC as the reference. Classes indicate reasons why testing was not performed. Dark green areas are masked due to dense vegetation.

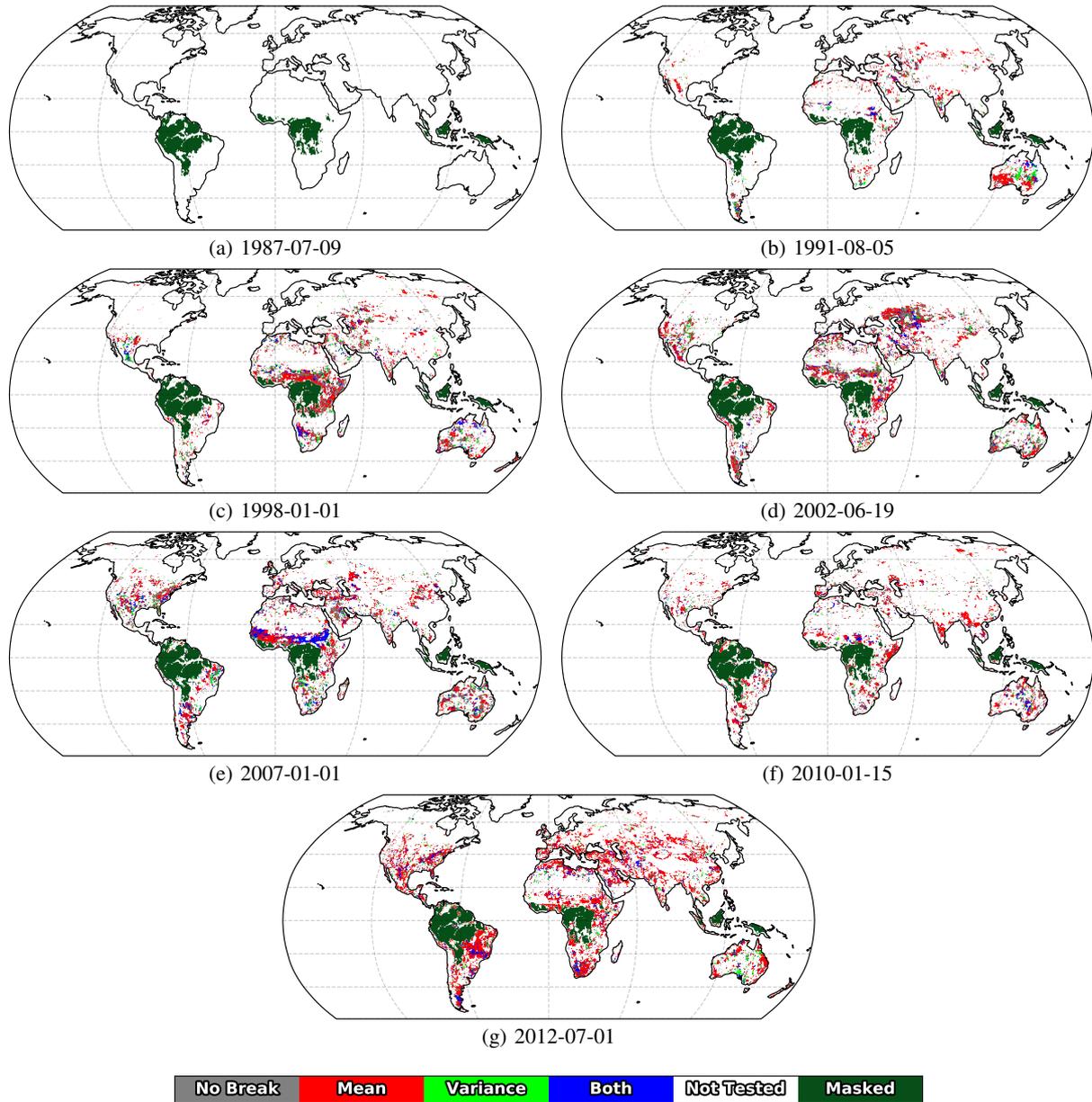


Fig. 9: Break testing with extended periods, before LMP adjustment with MERRA2 SFMC as the reference. The first transition date is never tested. Dark green areas are masked due to dense vegetation.

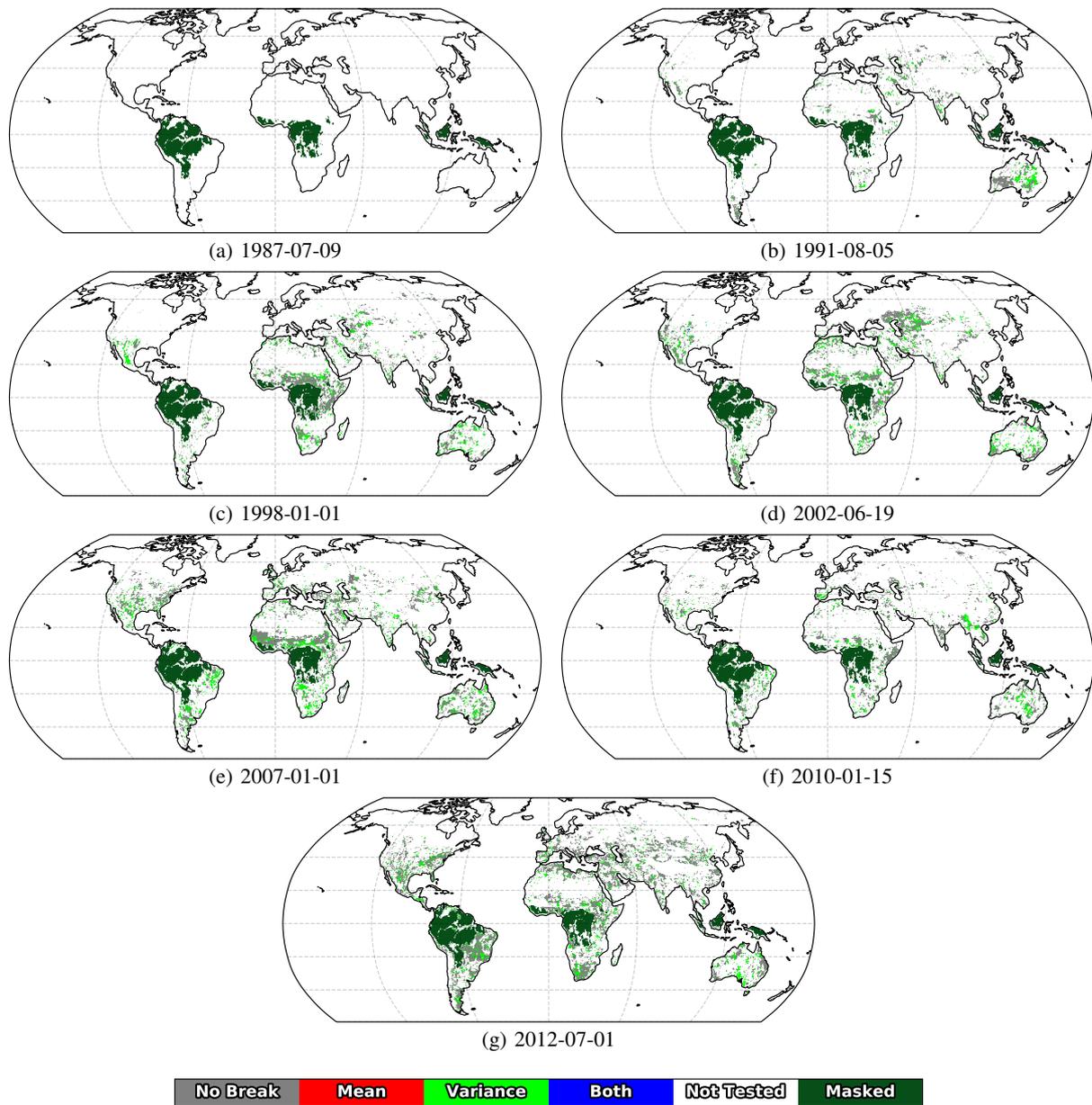


Fig. 10: Break testing with extended periods, after LMP adjustment with MERRA2 SFMC as the reference. The first transition date is never adjusted. Dark green areas are masked due to dense vegetation.

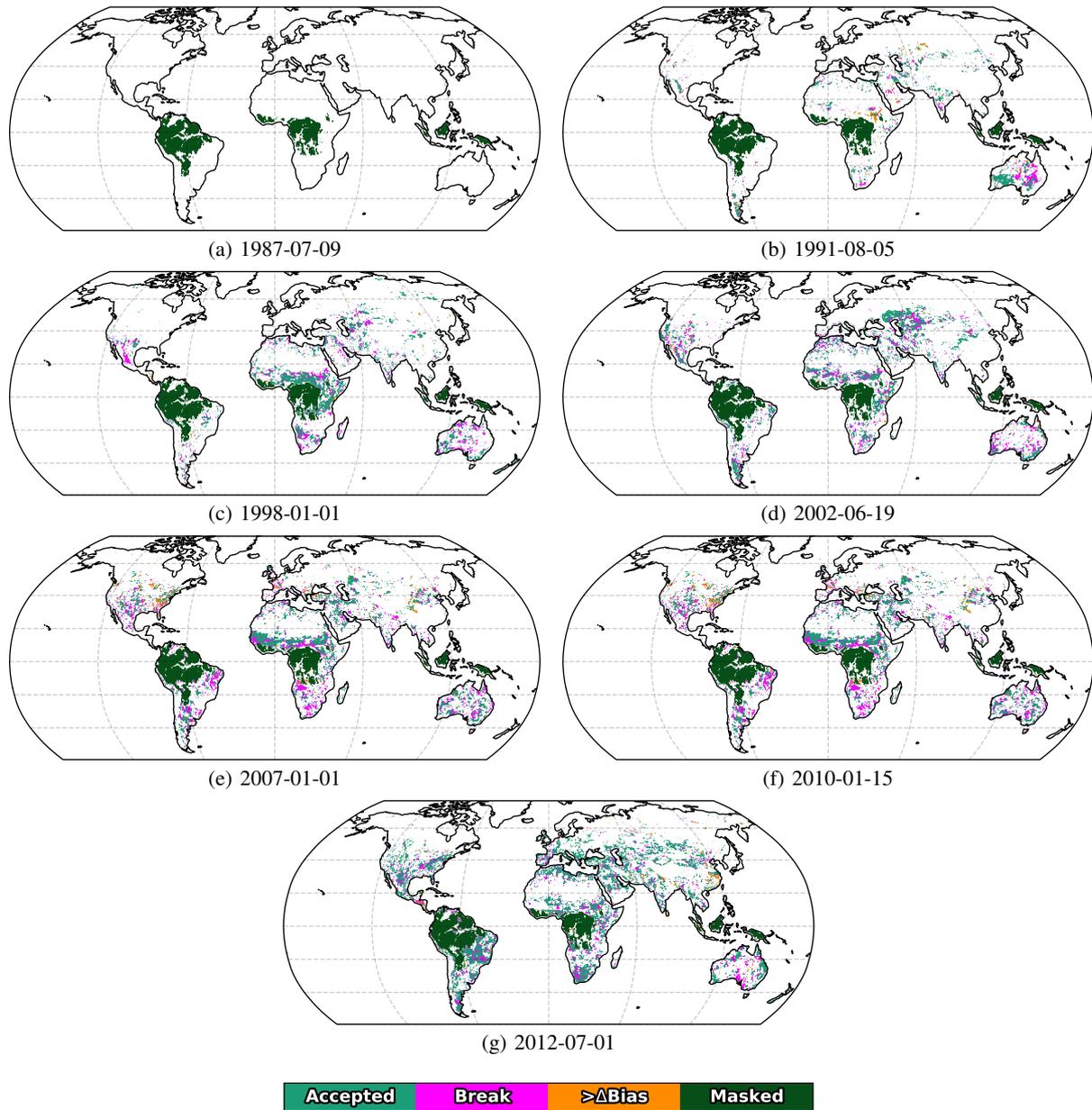


Fig. 11: Adjustment coverage (with extended periods) during LMP adjustment with ESA CCI SM v04.5 COMBINED and MERRA2 SFMC as the reference. The first transition date is never adjusted. Dark green areas are masked due to dense vegetation.

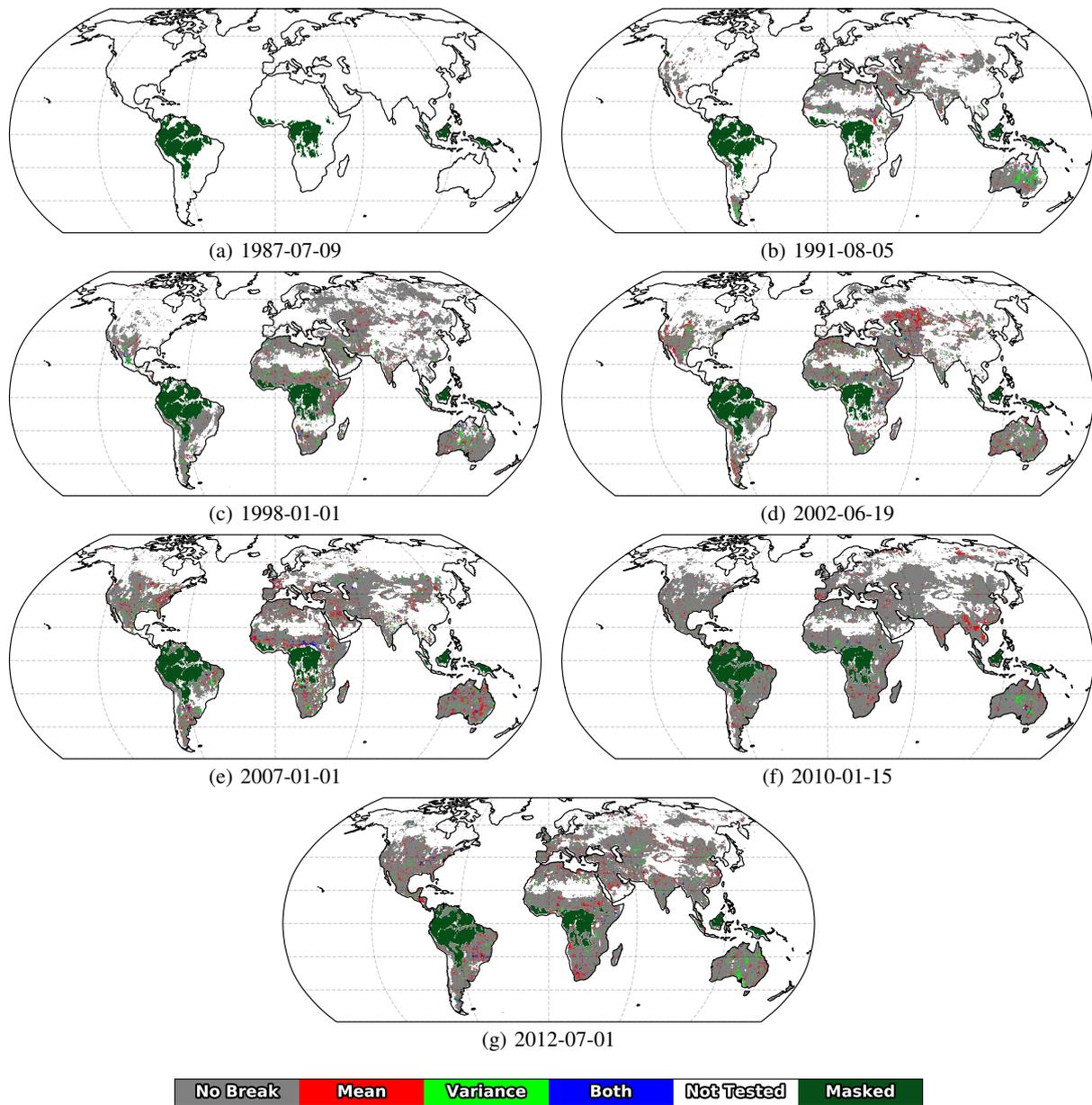


Fig. 12: Break testing after LMP adjustment with MERRA2 SFMC as the reference. The first transition date is never tested. Dark green areas are masked due to dense vegetation.

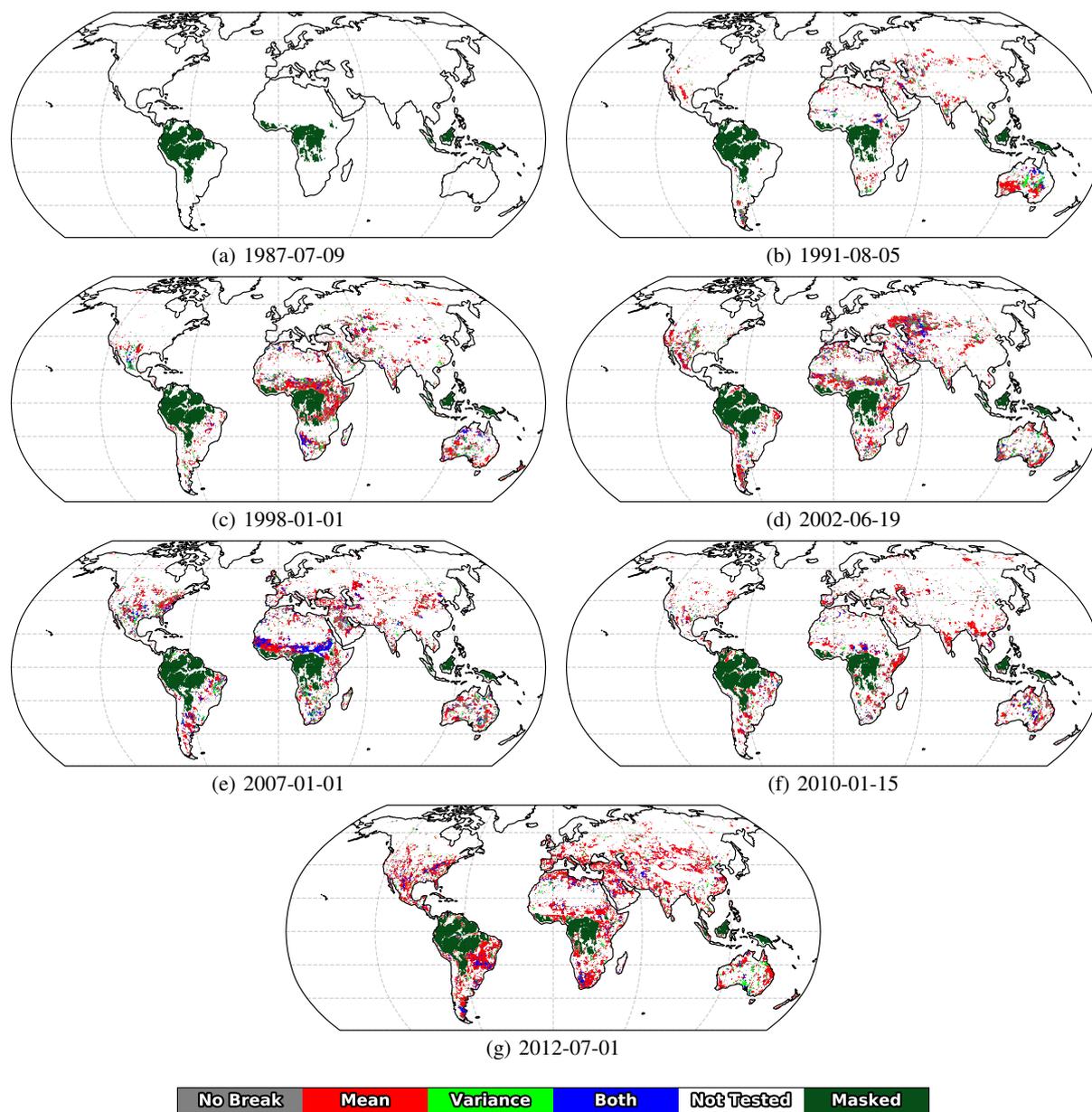


Fig. 13: Break testing with extended periods, before HOM adjustment with MERRA2 SFMC as the reference. The first transition date is never tested. Dark green areas are masked due to dense vegetation.

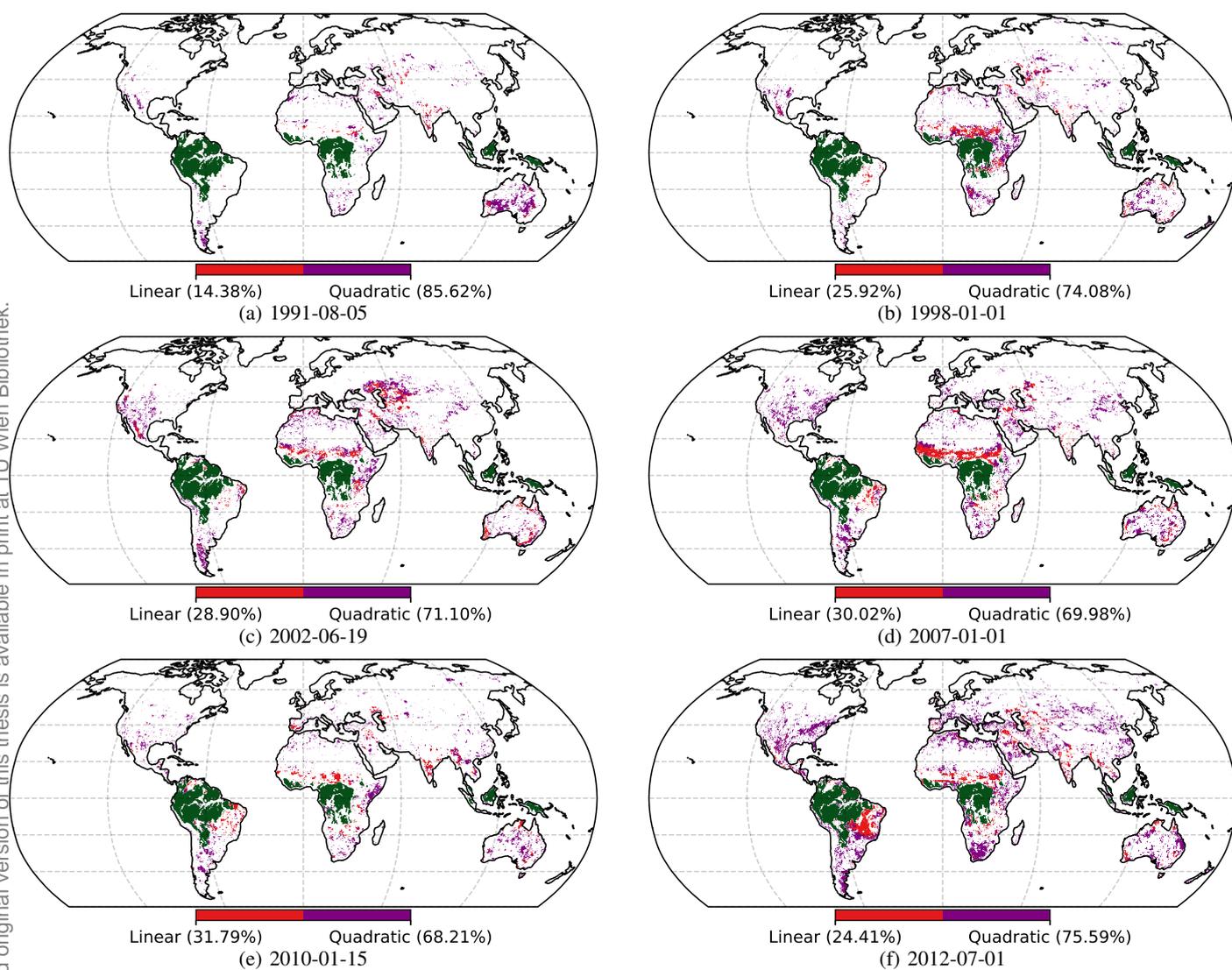


Fig. 14: Degree of the polynomial that is fitted during HOM adjustment, based on the threshold for using a linear model if $R_P > 0.8$. The model is then used to create predictions for adjustment.

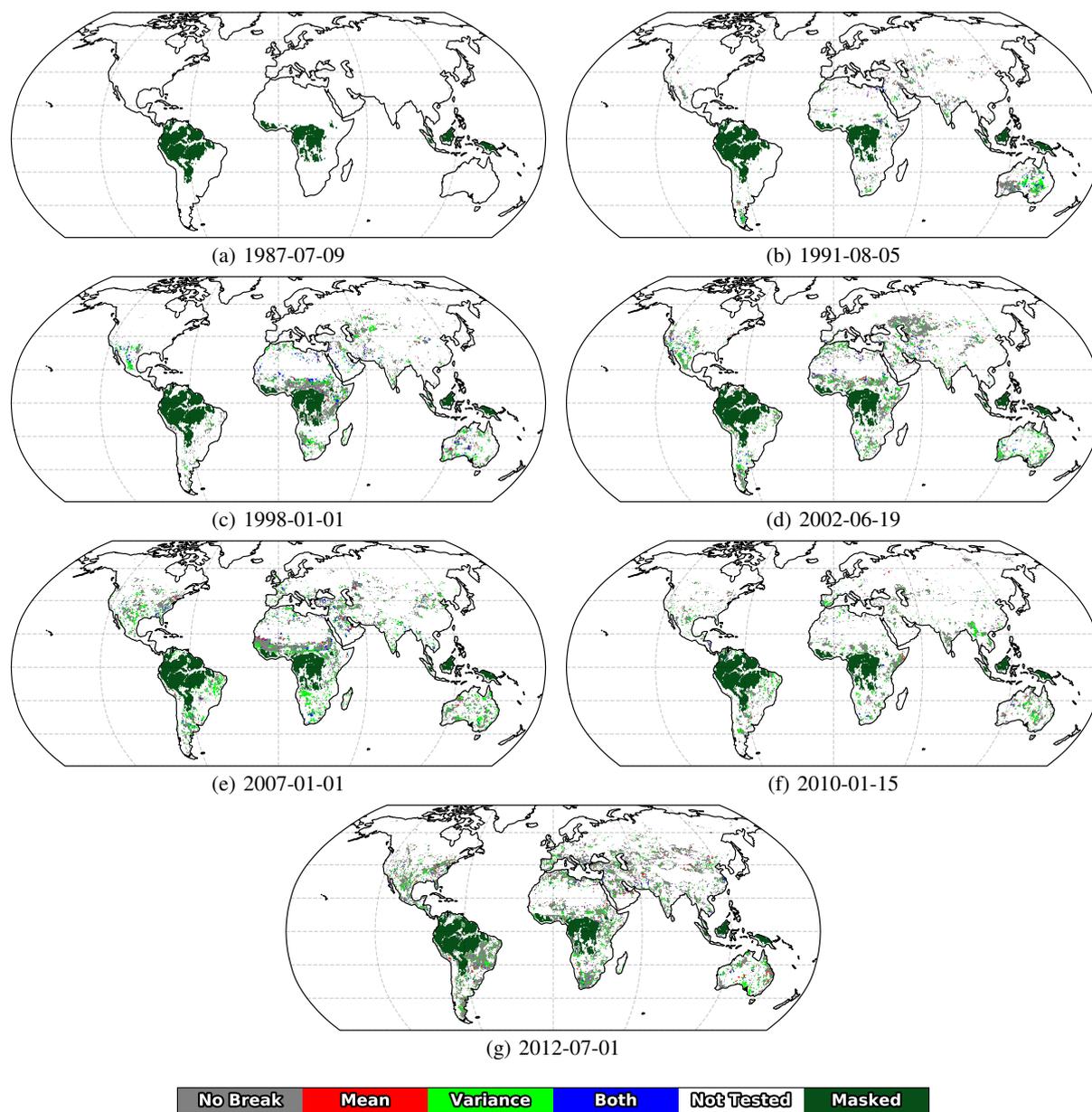


Fig. 15: Break testing with extended periods, after HOM adjustment with MERRA2 SFMC as the reference. The first transition date is never adjusted. Dark green areas are masked due to dense vegetation.

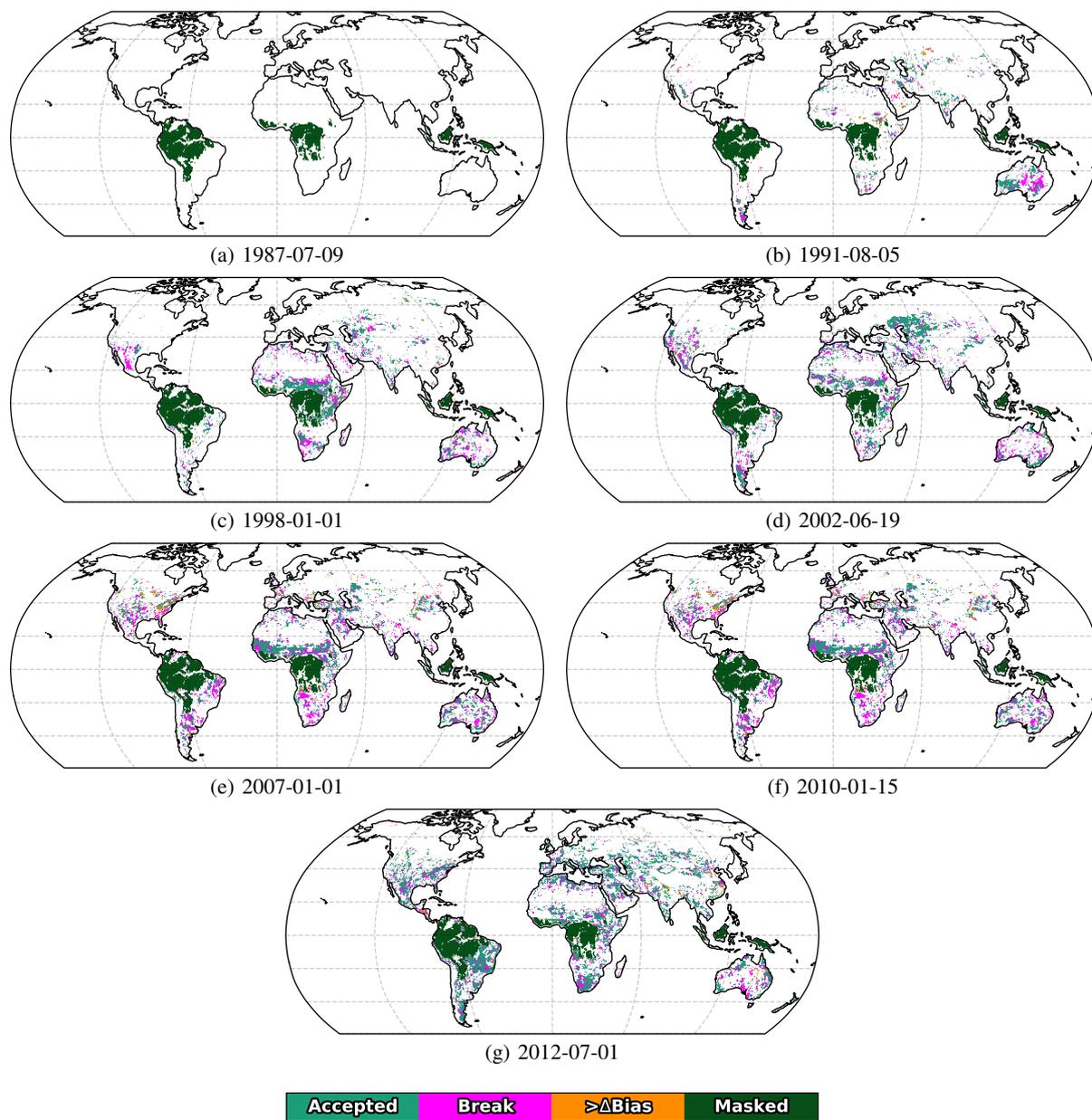


Fig. 16: Adjustment coverage (with extended periods) during HOM adjustment with ESA CCI SM v04.5 COMBINED and MERRA2 SFMC as the reference. The first transition date is never adjusted. Dark green areas are masked due to dense vegetation.

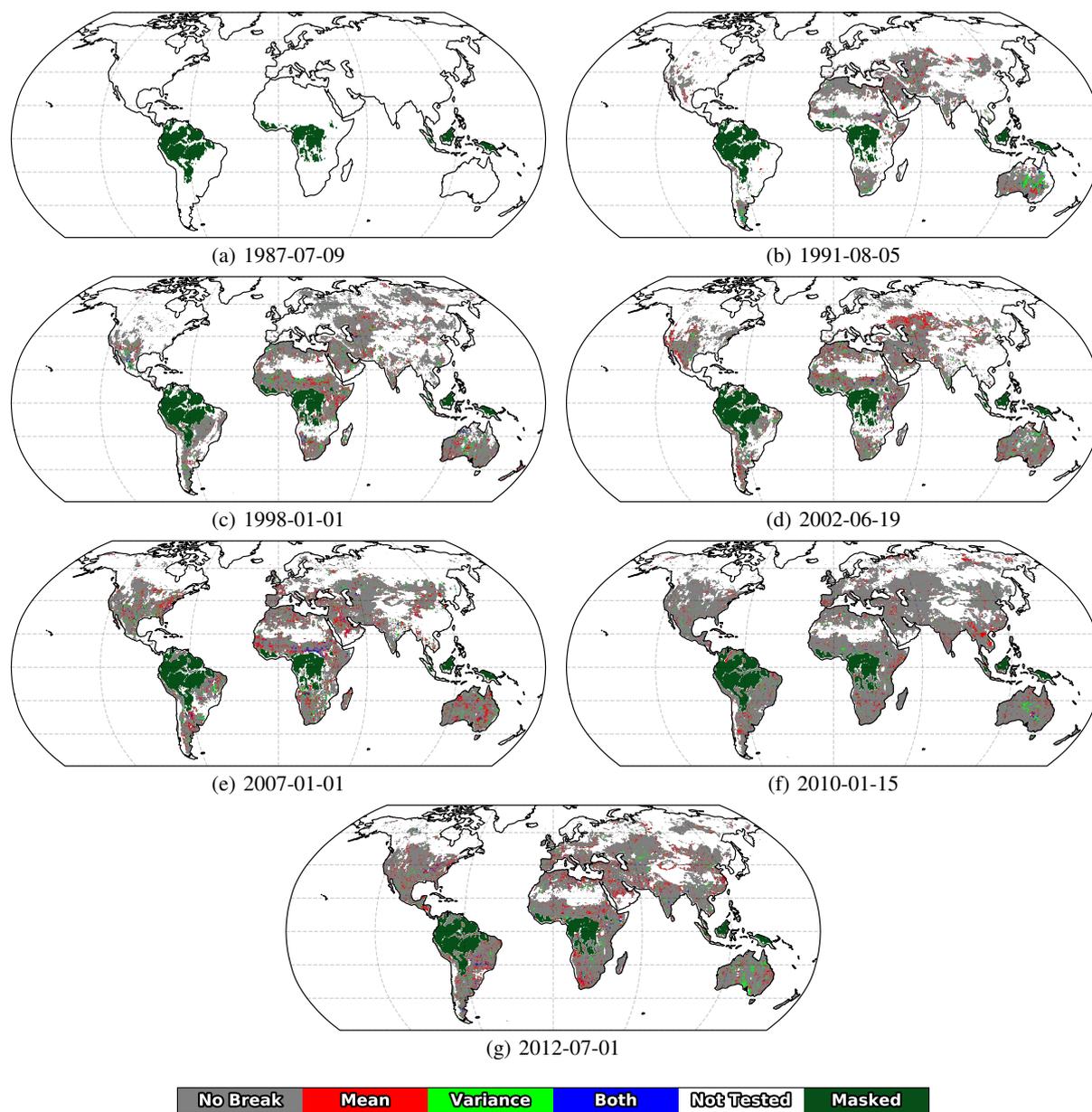


Fig. 17: Detected breaks in a separate test run after HOM adjustment with MERRA2 SFMC as the reference. The first transition date is never tested. Dark green areas are masked due to dense vegetation.

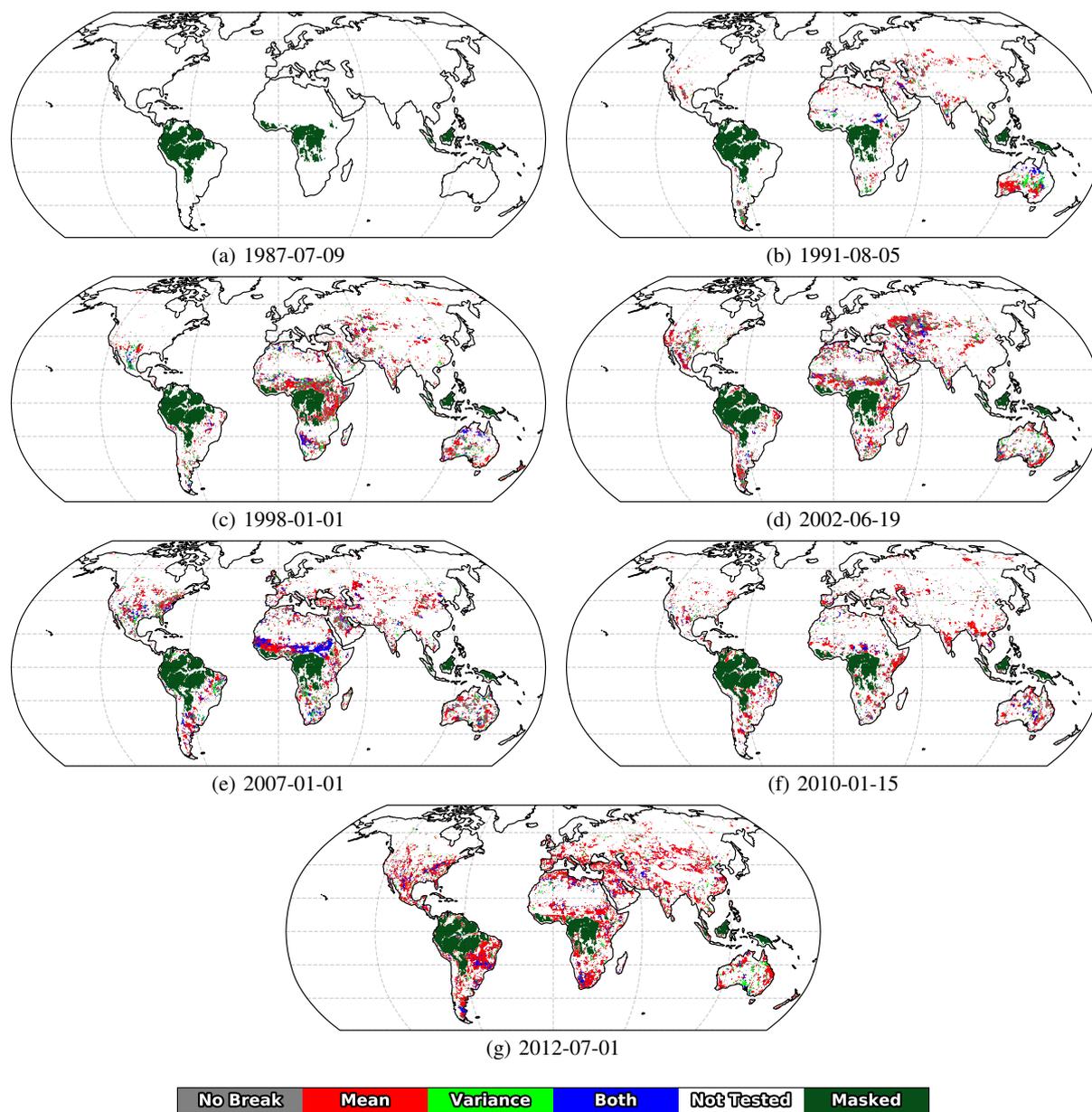


Fig. 18: Break testing with extended periods, before QCM adjustment with MERRA2 SFMC as the reference. The first transition date is never tested. Dark green areas are masked due to dense vegetation.

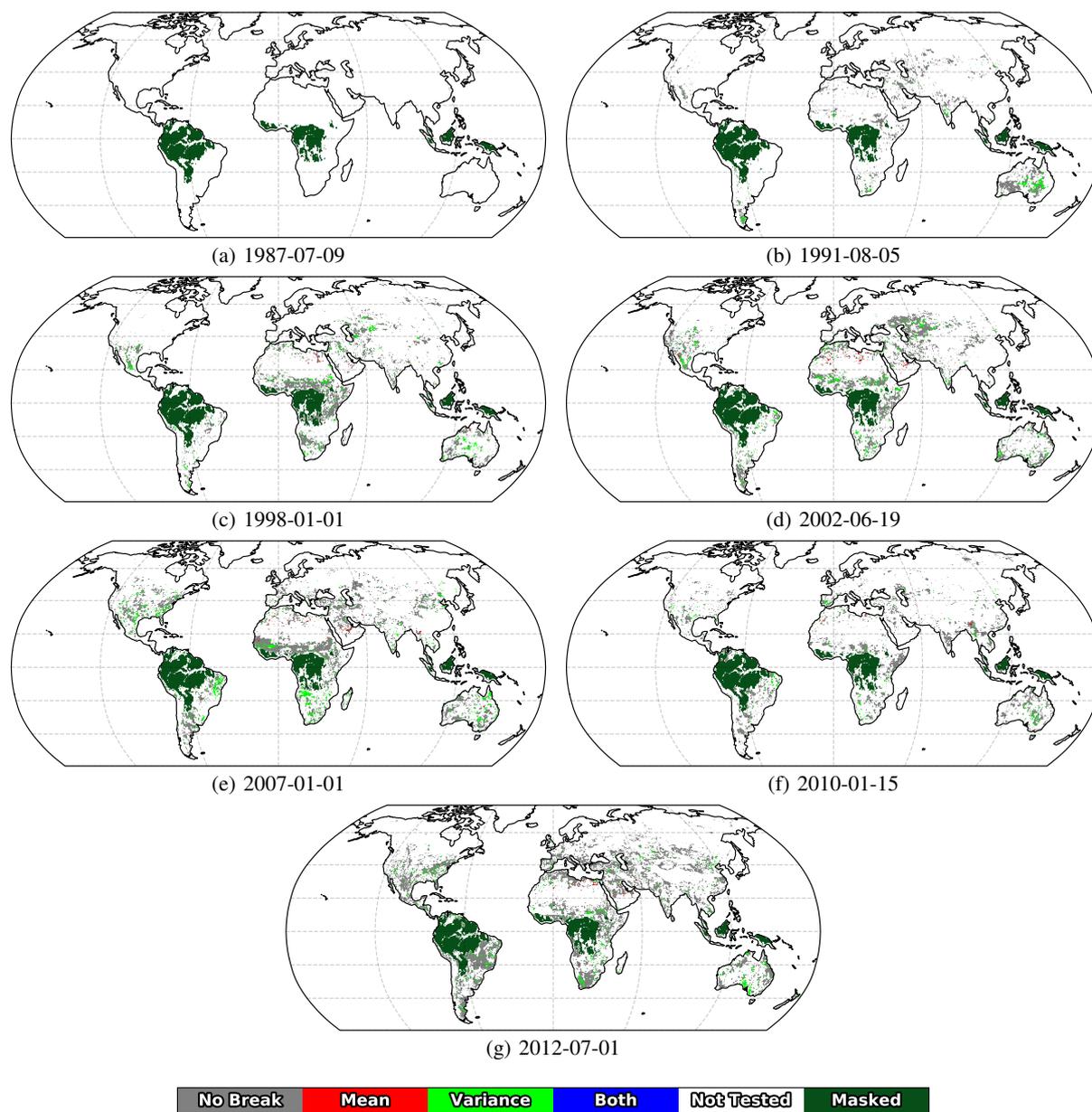


Fig. 19: Break testing with extended periods, after QCM adjustment with MERRA2 SFMC as the reference. The first transition date is never adjusted. Dark green areas are masked due to dense vegetation.

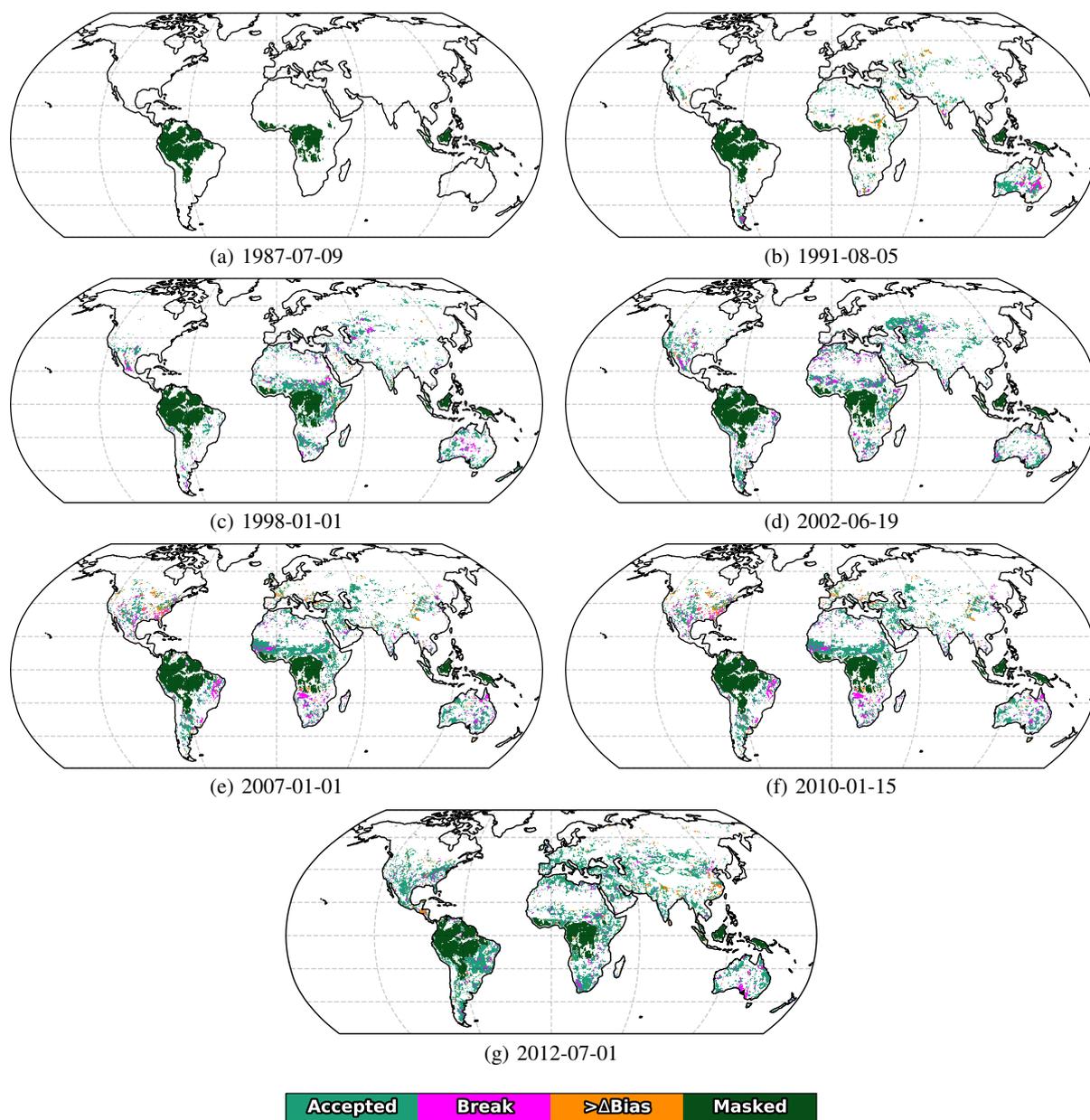


Fig. 20: Adjustment coverage (with extended periods) during QCM adjustment with ESA CCI SM v04.5 COMBINED and MERRA2 SFMC as the reference. The first transition date is never adjusted. Dark green areas are masked due to dense vegetation.

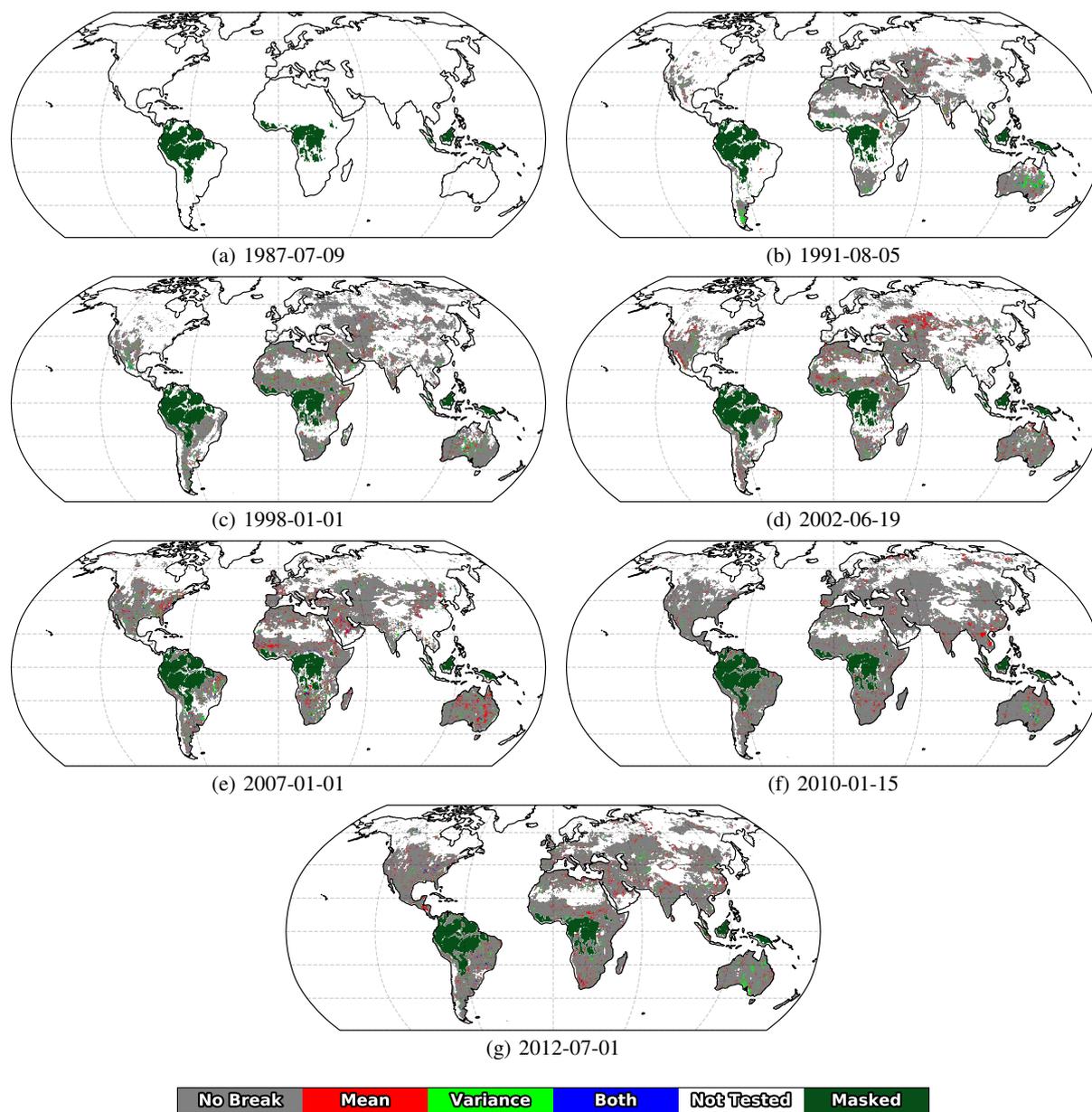


Fig. 21: Break testing after QCM adjustment with MERRA2 SFMC as the reference. The first transition date is never tested. Dark green areas are masked due to dense vegetation.

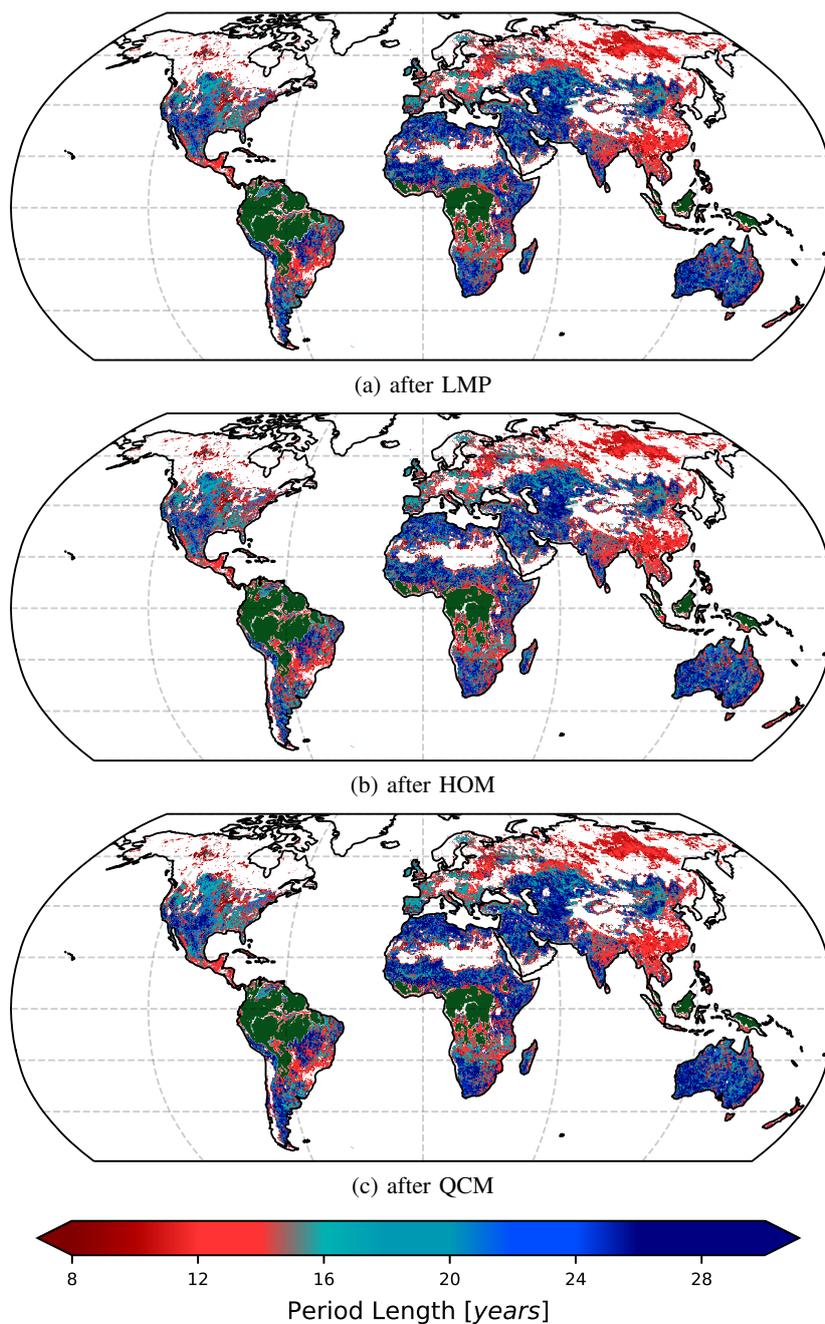


Fig. 22: Length of the longest homogeneous sub-period in the LMP, HOM and QCM adjusted ESA CCI SM v04.5 COMBINED dataset. Green areas indicate masking due to dense vegetation.

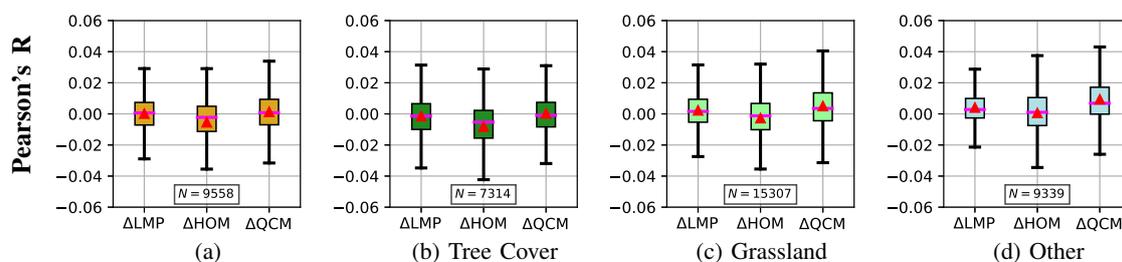


Fig. 23: Difference in Pearson's R between the unadjusted and adjusted versions of ESA CCI SM v04.5 COMBINED with ERA5 SWVL1 as the reference. Results are binned based on ESA CCI Land Cover classes. \blacktriangle indicates the mean of all values, $-$ the median, box edges represent the quartiles and whiskers range until $1.5 \cdot IQR$ of the difference values.

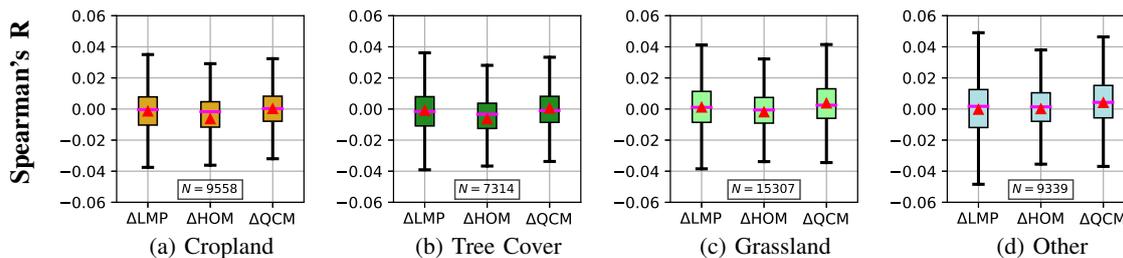


Fig. 24: Difference in Spearman's R between the unadjusted and adjusted versions of ESA CCI SM v04.5 COMBINED with ERA5 SWVL1 as the reference.

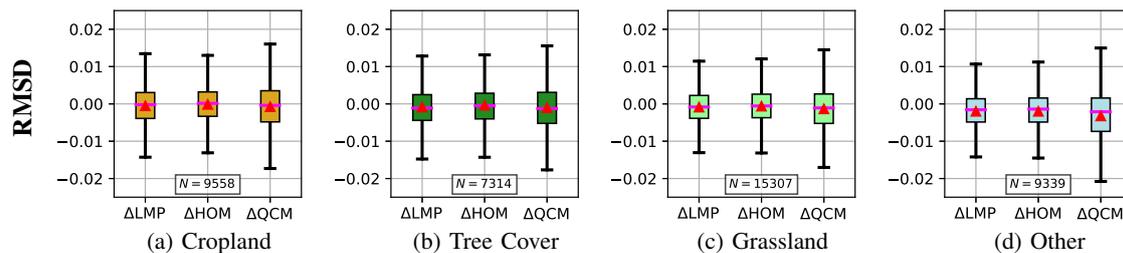


Fig. 25: Difference in RMSD between the unadjusted and adjusted versions of ESA CCI SM v04.5 COMBINED with ERA5 SWVL1 as the reference.

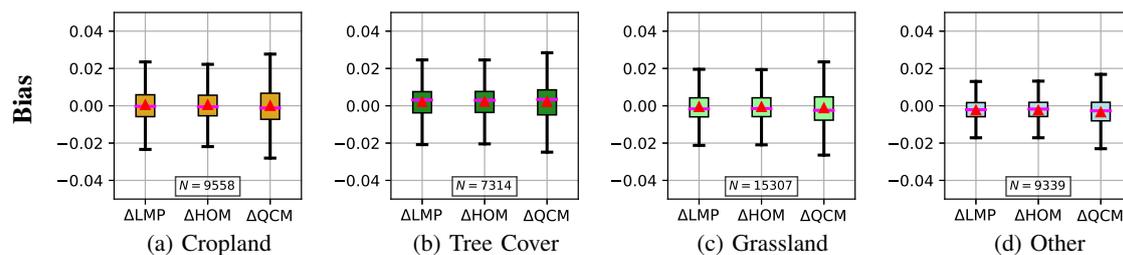


Fig. 26: Difference in Bias between the unadjusted and adjusted versions of ESA CCI SM v04.5 COMBINED with ERA5 SWVL1 as the reference.

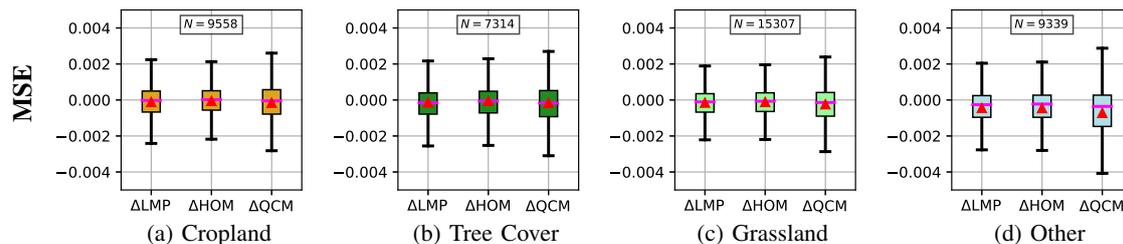


Fig. 27: Difference in MSE between the unadjusted and adjusted versions of ESA CCI SM v04.5 COMBINED with ERA5 SWVL1 as the reference.

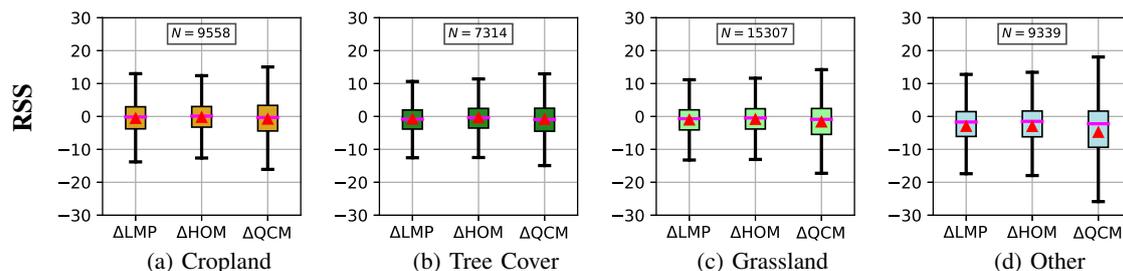


Fig. 28: Difference in RSS between the unadjusted and adjusted versions of ESA CCI SM v04.5 COMBINED with ERA5 SWVL1 as the reference.

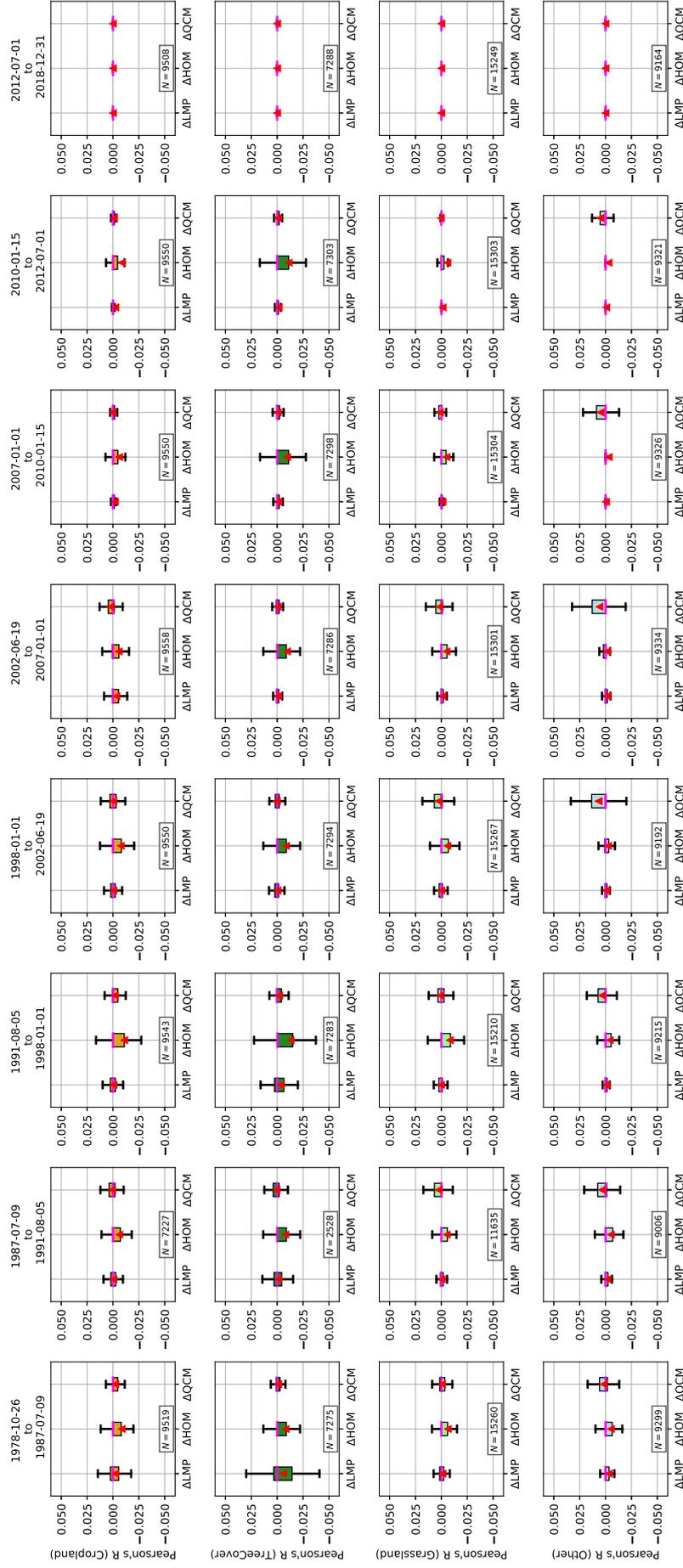


Fig. 29: Intercomparison of Pearson's R across different land cover classes, with ERA5 SWVL1 as the reference. Differences between the unadjusted and adjusted (LMP, HOM, QCM) data sets. \blacktriangle indicates the mean of all values, $—$ the median, box edges represent the quartiles and whiskers range until $1.5 * IQR$ of the difference values.

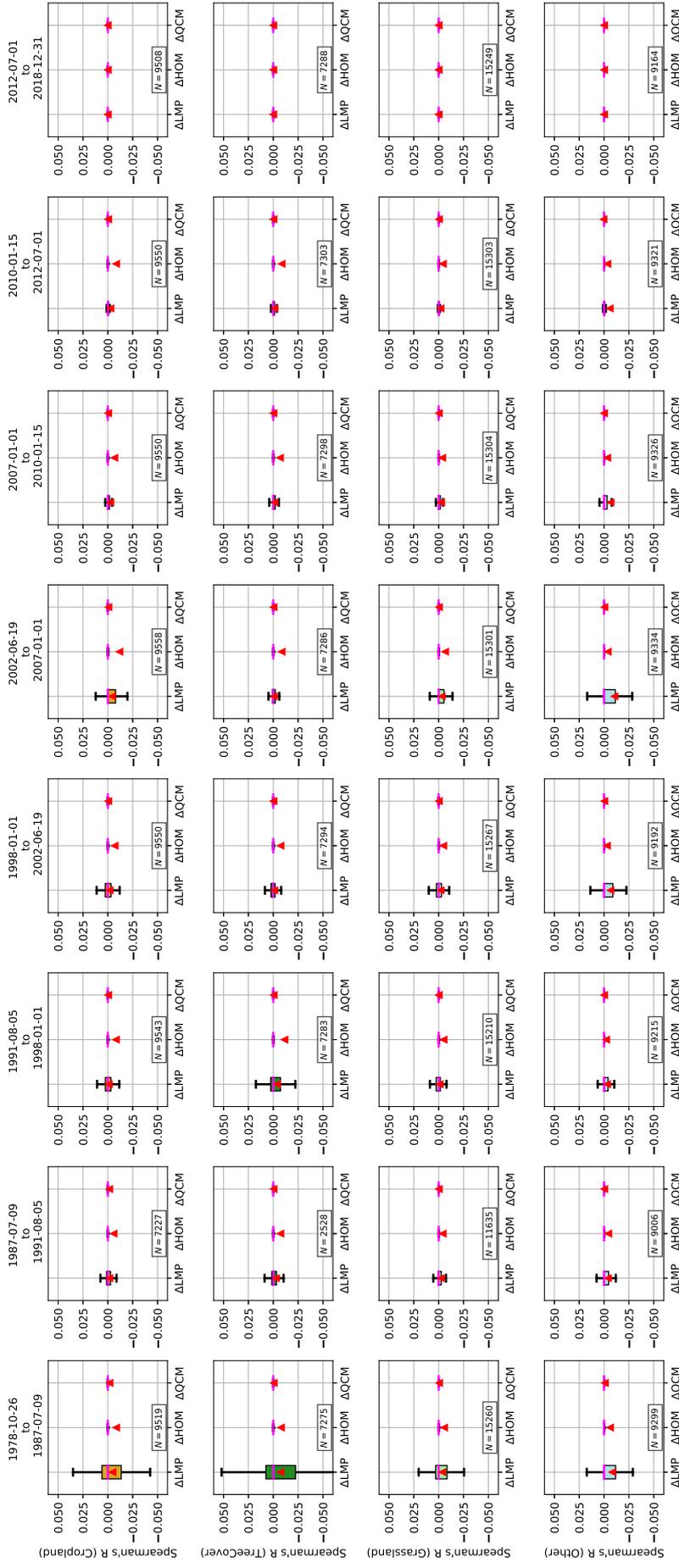


Fig. 30: Intercomparison of Spearman's R across different land cover classes, with ERA5 SWVL1 as the reference. Differences between the unadjusted and adjusted (LMP, HOM, QCM) data sets. \blacktriangle indicates the mean of all values, $—$ the median, box edges represent the quartiles and whiskers range until $1.5 * IQR$ of the difference values.

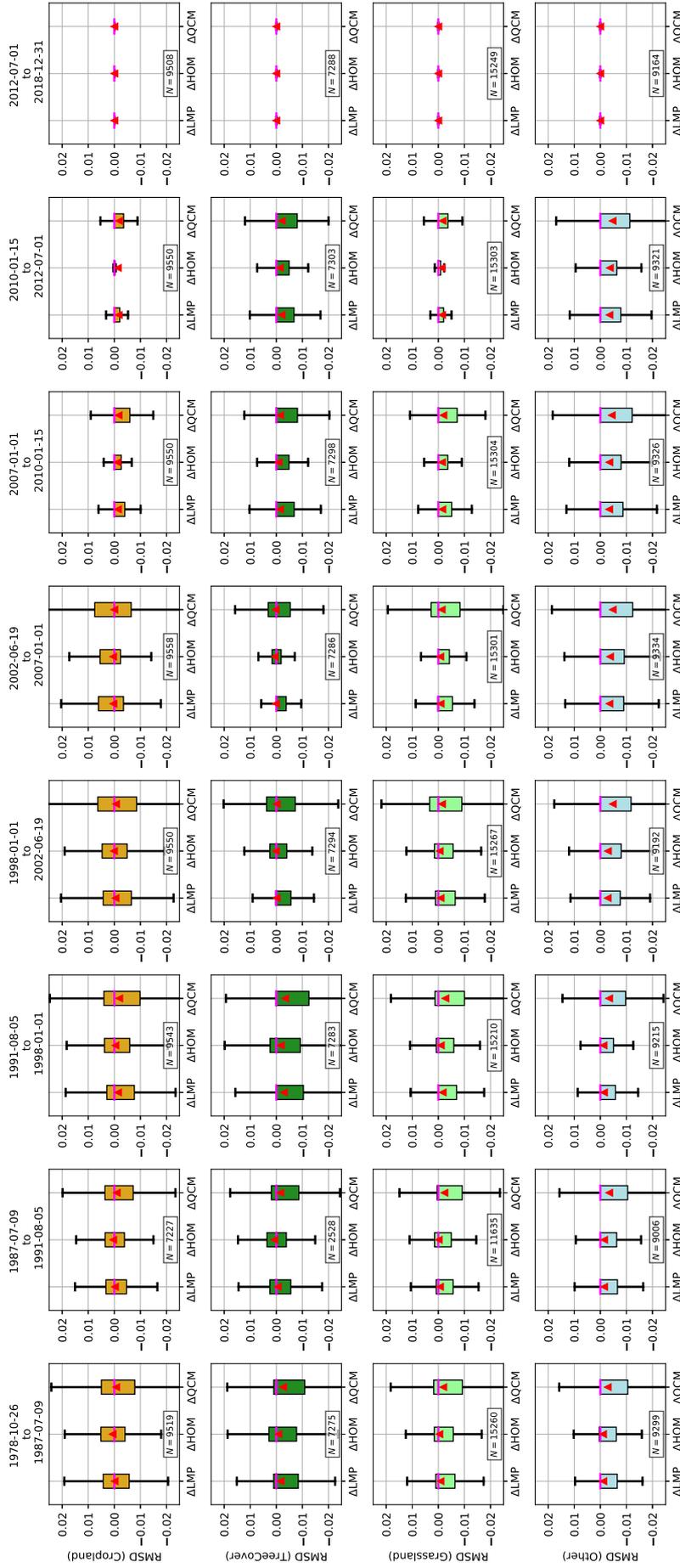


Fig. 31: Intercomparison of RMSD across different land cover classes, with ERA5 SWVL1 as the reference. Differences between the unadjusted and adjusted (LMP, HOM, QCM) data sets. \blacktriangle indicates the mean of all values, $—$ the median, box edges represent the quartiles and whiskers range until $1.5 \cdot IQR$ of the difference values.

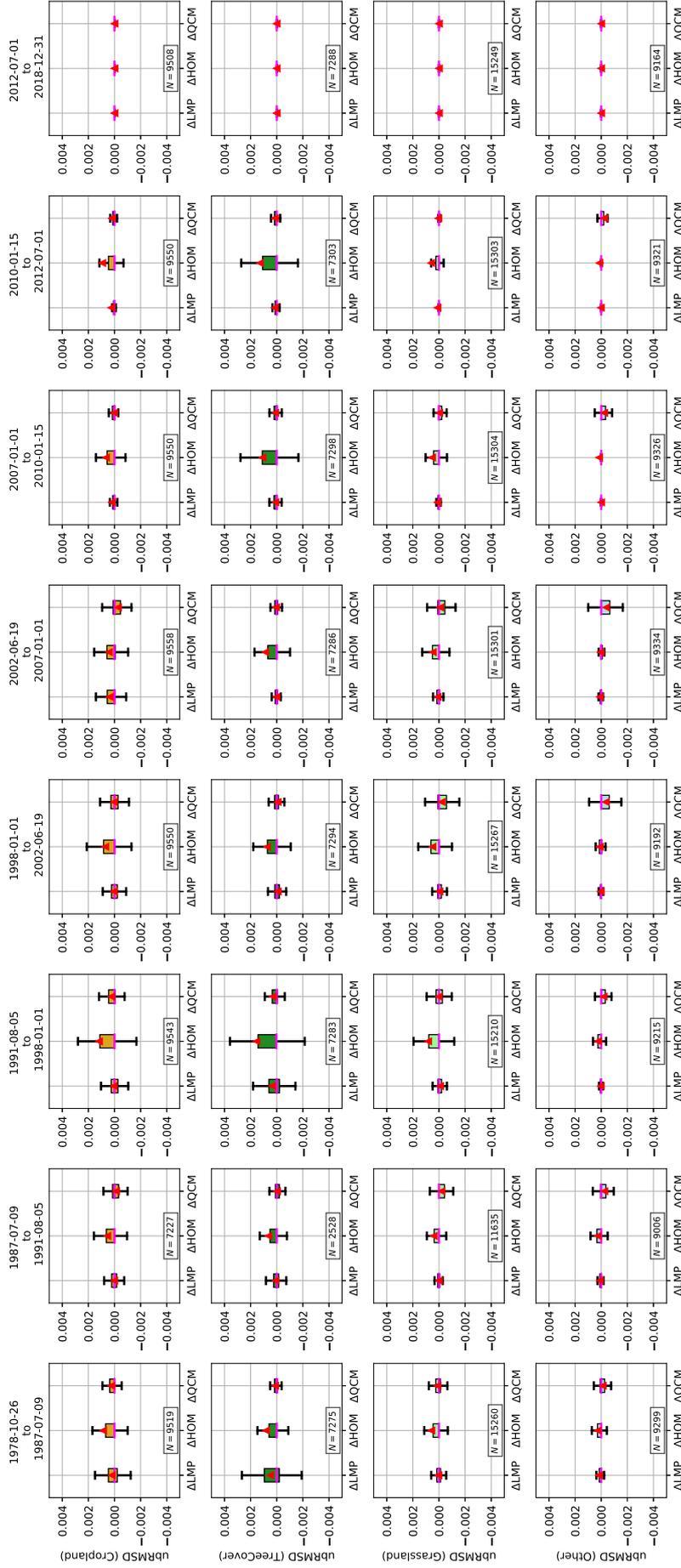


Fig. 32: Intercomparison of ubRMSD across different land cover classes, with ERA5 SWVL1 as the reference. Differences between the unadjusted and adjusted (LMP, HOM, QCM) data sets. \blacktriangle indicates the mean of all values, $—$ the median, box edges represent the quartiles and whiskers range until $1.5 * IQR$ of the difference values.

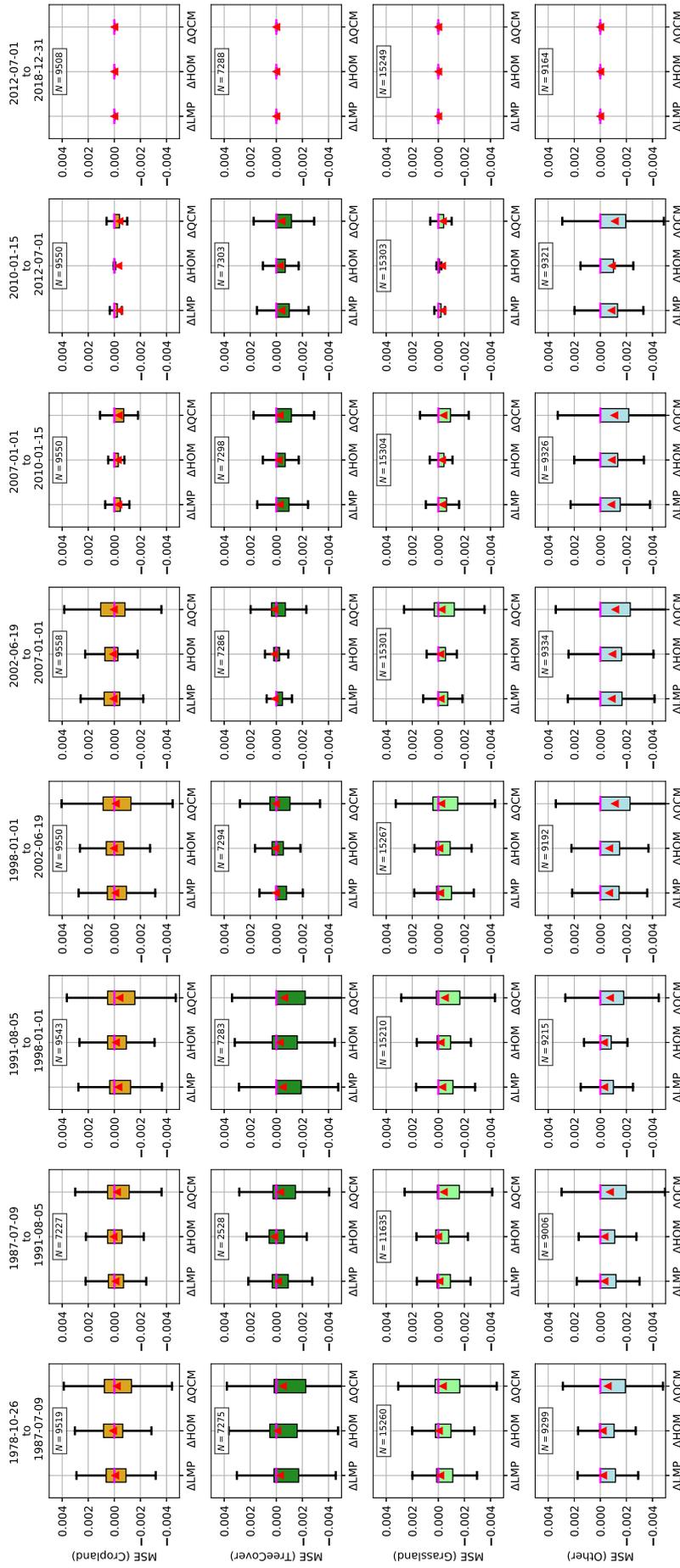


Fig. 33: Intercomparison of MSE across different land cover classes, with ERA5 SWVL1 as the reference. Differences between the unadjusted and adjusted (LMP, HOM, QCM) data sets. \blacktriangle indicates the mean of all values, $—$ indicates the median, box edges represent the quartiles and whiskers range until $1.5 \cdot IQR$ of the difference values.

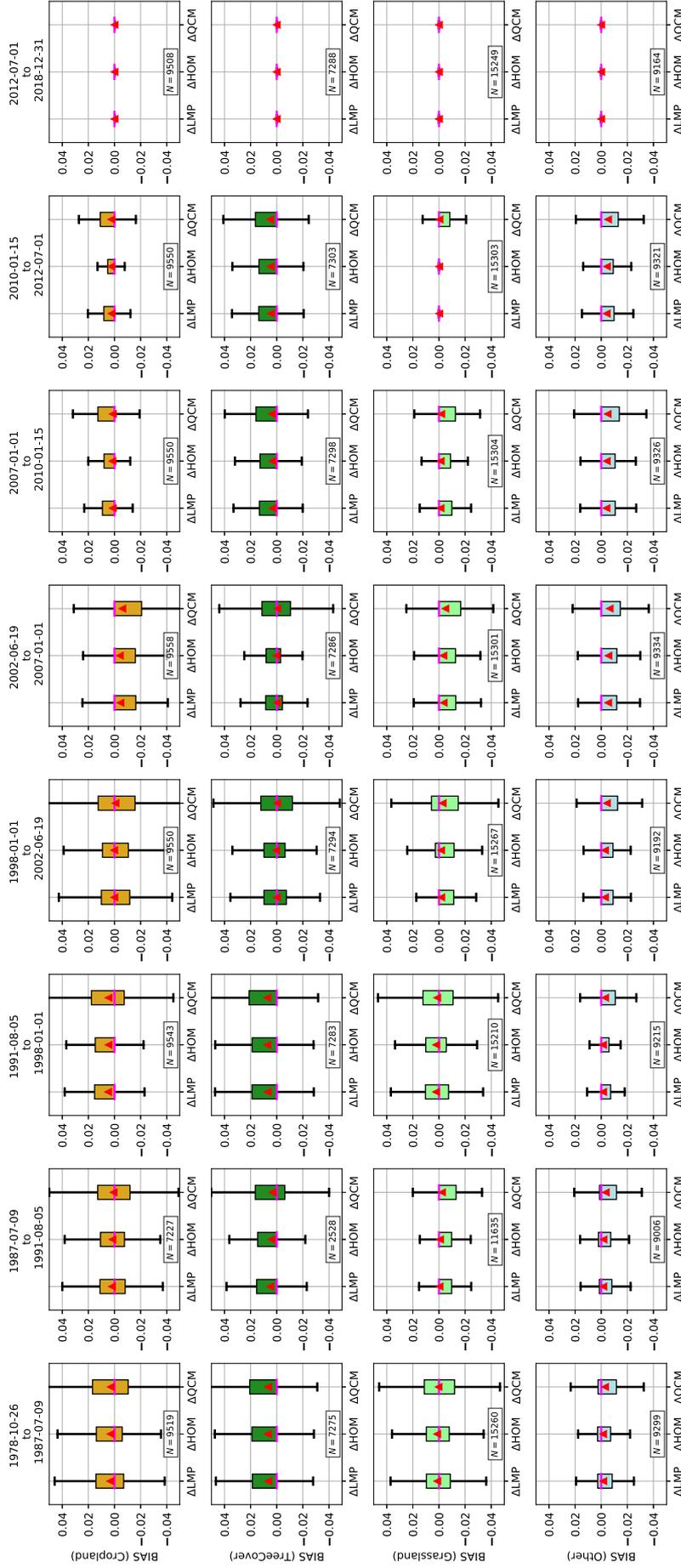


Fig. 34: Intercomparison of Bias across different land cover classes, with ERA5 SWVL1 as the reference. Differences between the unadjusted and adjusted (LMP, AHOM, QCM) data sets. \blacktriangle indicates the mean of all values, $—$ the median, box edges represent the quartiles and whiskers range until $1.5 \cdot IQR$ of the difference values.

| | | 1978-10-26 to 1987-07-09 | 1987-07-09 to 1991-08-05 | 1991-08-05 to 1998-01-01 | 1998-01-01 to 2002-06-19 | 2002-06-19 to 2007-01-01 | 2007-01-01 to 2010-01-15 | 2010-01-15 to 2012-07-01 | 2012-07-01 to 2018-06-30 |
|----------------|-------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|
| ΔR_P | LMP | -2.63E-03 | -2.43E-03 | -1.30E-05 | 8.49E-04 | -1.69E-03 | -2.34E-03 | -2.03E-03 | 0.00E+00 |
| | HOM | -7.44E-03 | -5.10E-03 | -7.81E-03 | -4.01E-03 | -2.81E-03 | -3.97E-03 | -3.73E-03 | 0.00E+00 |
| | QCM | 2.35E-04 | 4.04E-03 | 6.87E-04 | 3.65E-03 | 3.12E-03 | 2.02E-03 | 1.52E-03 | -2.74E-06 |
| Δ RMSD | LMP | -1.62E-03 | -1.24E-03 | -2.80E-03 | -1.37E-03 | -8.37E-04 | -2.06E-03 | -2.20E-03 | 6.63E-07 |
| | HOM | -1.13E-03 | -9.76E-04 | -2.09E-03 | -1.08E-03 | -6.53E-04 | -1.68E-03 | -1.68E-03 | 0.00E+00 |
| | QCM | -2.66E-03 | -2.88E-03 | -3.88E-03 | -1.85E-03 | -1.15E-03 | -2.43E-03 | -2.59E-03 | 2.65E-06 |
| Δ RSS | LMP | -1.15E-01 | -3.90E-02 | -3.18E-01 | -6.20E-02 | -1.01E-03 | -2.39E-01 | -2.10E-01 | 1.04E-03 |
| | HOM | -8.04E-02 | -1.64E-02 | -2.19E-01 | -3.73E-02 | 2.64E-02 | -1.89E-01 | -1.71E-01 | 1.07E-03 |
| | QCM | -1.82E-01 | -1.24E-01 | -4.33E-01 | -1.10E-01 | -4.51E-02 | -2.71E-01 | -2.45E-01 | 1.07E-03 |
| Δ mse | LMP | -3.11E-04 | -2.34E-04 | -4.96E-04 | -2.39E-04 | -1.40E-04 | -3.29E-04 | -3.52E-04 | 1.02E-07 |
| | HOM | -2.17E-04 | -1.84E-04 | -3.71E-04 | -1.89E-04 | -1.09E-04 | -2.68E-04 | -2.70E-04 | 0.00E+00 |
| | QCM | -5.09E-04 | -5.38E-04 | -6.83E-04 | -3.23E-04 | -1.91E-04 | -3.87E-04 | -4.14E-04 | 4.12E-07 |
| ΔR_S | LMP | -5.48E-03 | -4.21E-03 | -1.54E-03 | -3.06E-03 | -3.44E-03 | -3.51E-03 | -3.53E-03 | 7.81E-06 |
| | HOM | -7.53E-03 | -4.82E-03 | -6.07E-03 | -4.41E-03 | -7.08E-03 | -4.00E-03 | -4.56E-03 | 0.00E+00 |
| | QCM | -9.73E-04 | -5.49E-04 | -4.08E-04 | -3.15E-04 | -3.54E-04 | -2.65E-04 | -3.28E-04 | 0.00E+00 |
| Δ urmsd | LMP | 2.52E-04 | -5.03E-05 | 5.07E-05 | 1.03E-06 | 1.94E-04 | 1.18E-04 | 1.08E-04 | 6.34E-06 |
| | HOM | 8.28E-04 | 4.00E-04 | 8.48E-04 | 4.52E-04 | 3.85E-04 | 6.08E-04 | 5.20E-04 | 7.25E-06 |
| | QCM | 1.56E-04 | -2.67E-04 | 1.33E-04 | -1.19E-04 | -2.91E-04 | 4.65E-05 | 4.09E-05 | 8.01E-06 |
| Avg. N Obs. | 339.2 | 399.1 | 784.7 | 666.5 | 790.6 | 757.8 | 626.0 | 1787.4 | |

TABLE I: Difference in (global) median of error metrics between the un-adjusted and all three adjusted (LMP, HOM, QCM) ESA CCI SM products compared to ERA5 SWVLI. The last line describes the average number of observations for all points that were intercompared (that were adjusted by all three methods). Green numbers indicate an improvement compared to ESA CCI SM before the correction.

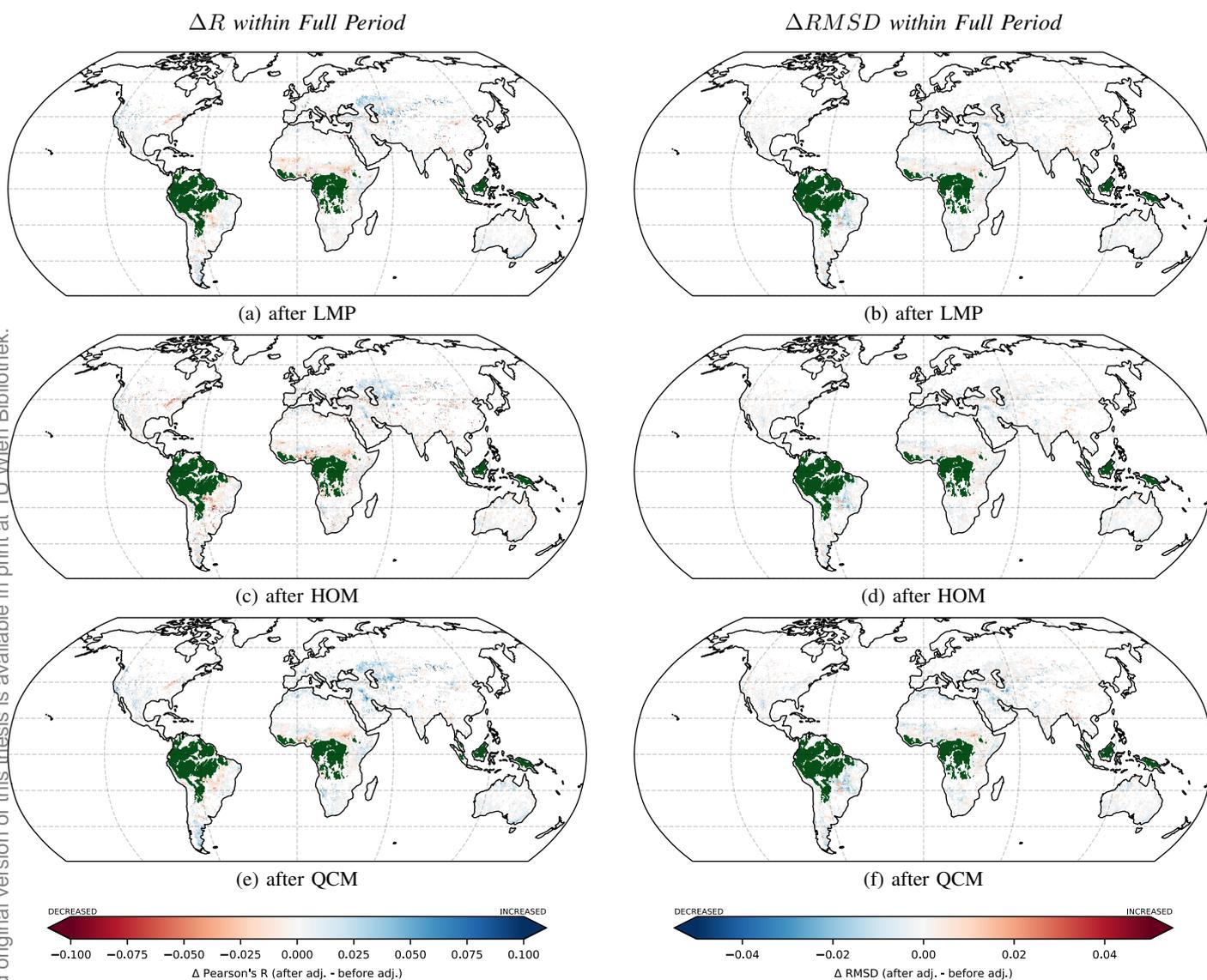
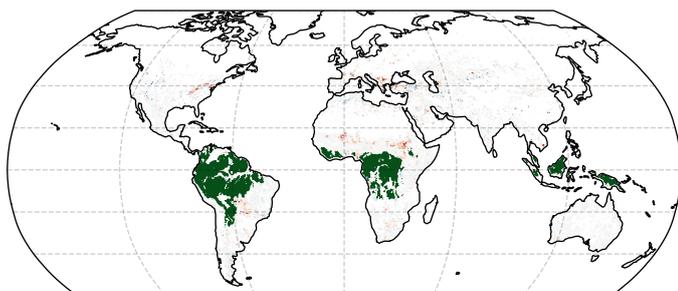
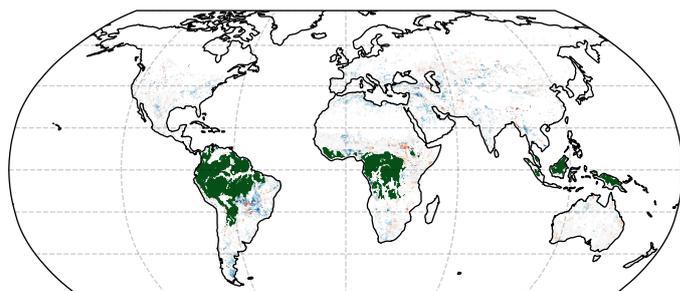


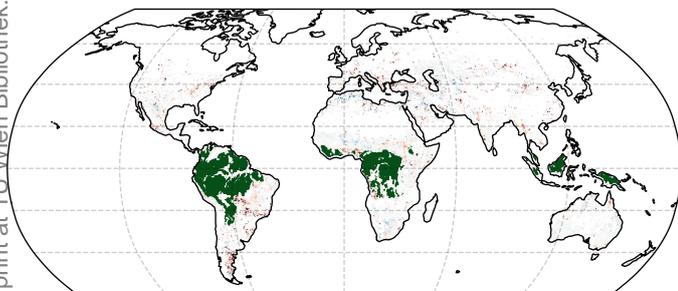
Fig. 35: Differences in Pearson's R (left) and RMSD (right) between the unadjusted and the adjusted versions of ESA CCI SM within the full ESA CCI SM v04.5 COMBINED period (1978-10-26 to 2018-12-31) with ERA5 SWVL1 as the reference.

ΔR in SSP_1 

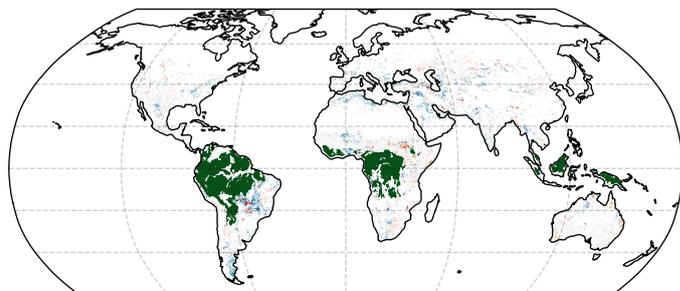
(a) after LMP

 $\Delta RMSD$ in SSP_1 

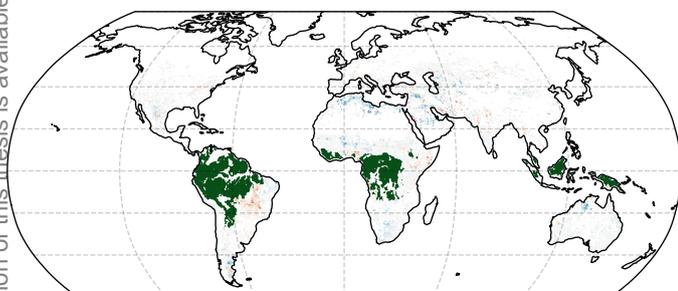
(b) after LMP



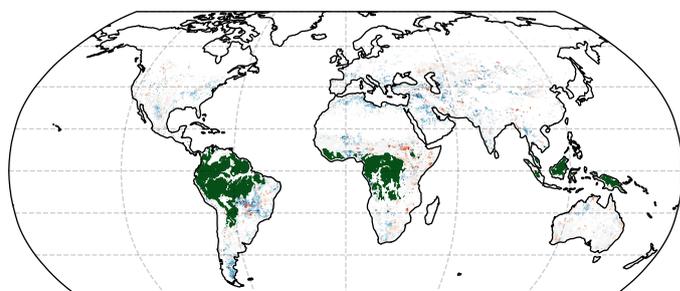
(c) after HOM



(d) after HOM



(e) after QCM



(f) after QCM

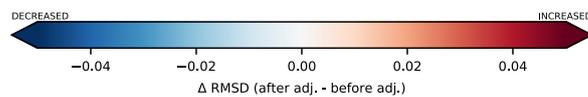
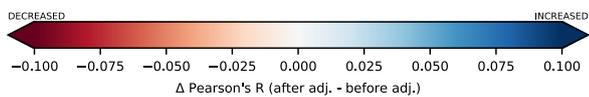
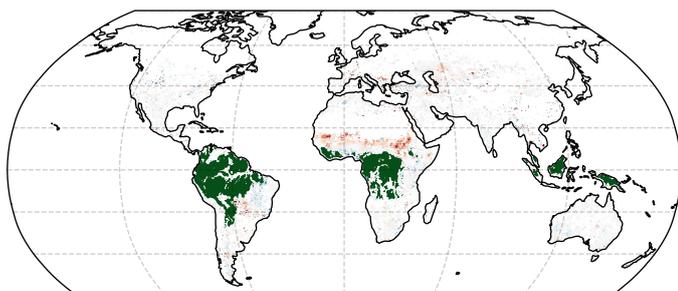
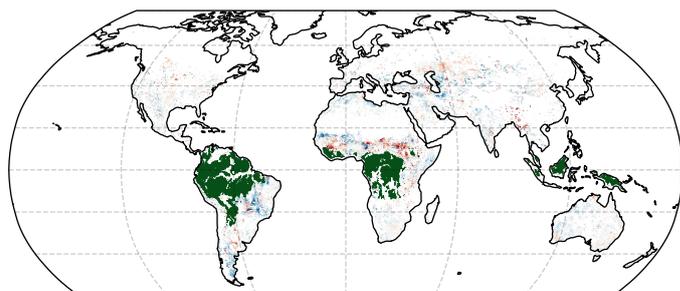


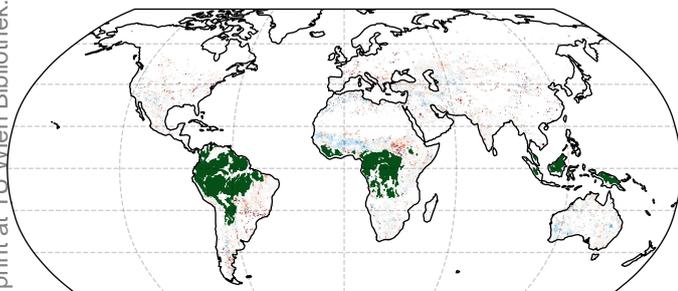
Fig. 36: Differences in Pearson's R (left) and RMSD (right) between the unadjusted and the adjusted versions of ESA CCI SM v04.5 COMBINED within SSP_1 (2010-01-15 to 2012-08-01) with ERA5 SWVL1 as the reference.

ΔR in SSP_3 

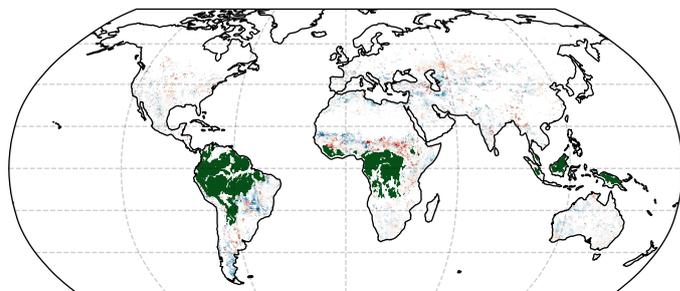
(a) after LMP

 $\Delta RMSD$ in SSP_3 

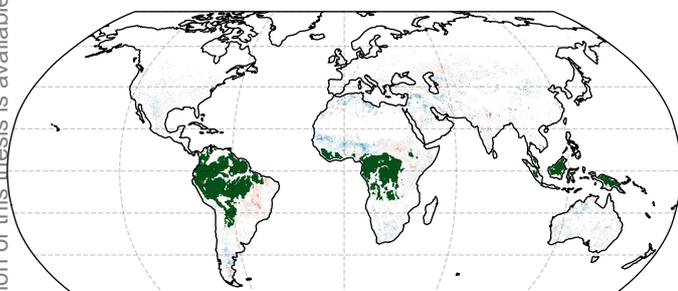
(b) after LMP



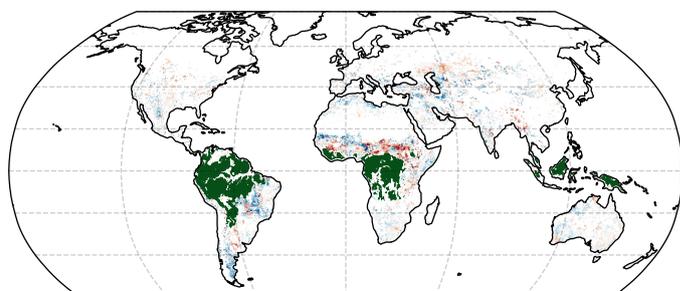
(c) after HOM



(d) after HOM



(e) after QCM



(f) after QCM

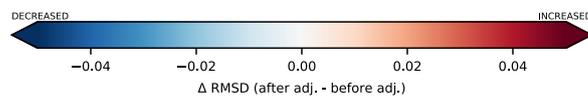
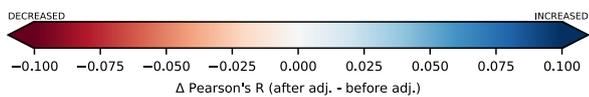
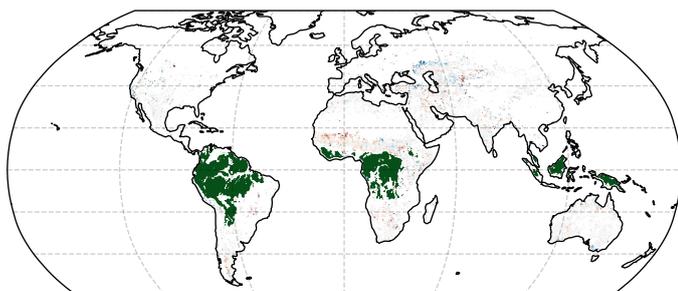
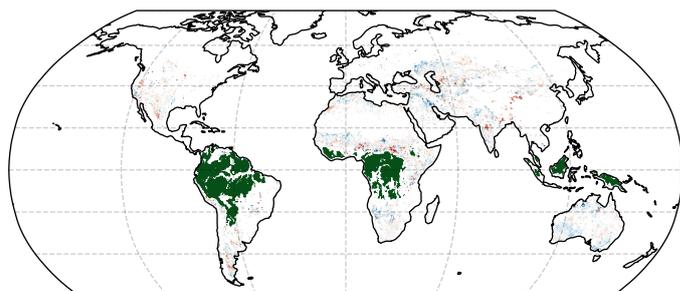


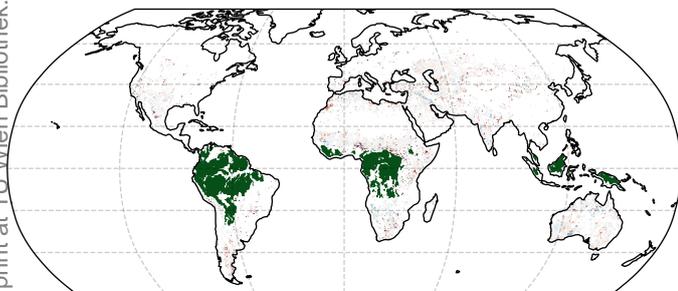
Fig. 37: Differences in Pearson's R (left) and RMSD (right) between the unadjusted and the adjusted versions of ESA CCI SM v04.5 COMBINED within SSP_3 (2002-06-19 to 2007-01-01) with ERA5 SWVL1 as the reference.

ΔR in SSP_6 

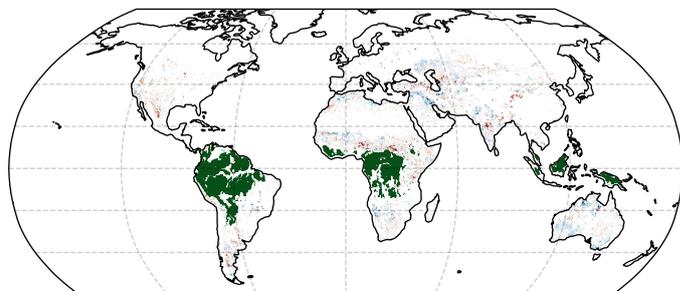
(a) after LMP

 $\Delta RMSD$ in SSP_6 

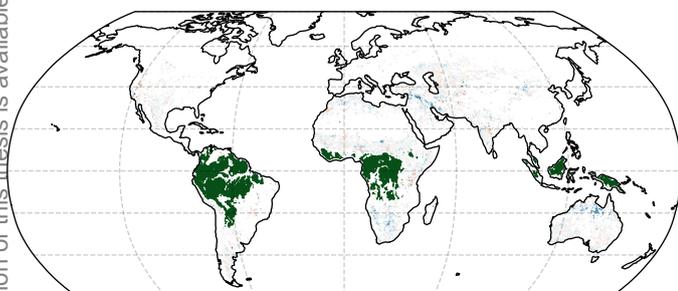
(b) after LMP



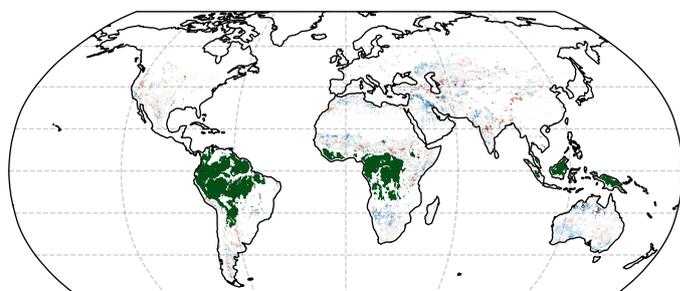
(c) after HOM



(d) after HOM



(e) after QCM



(f) after QCM

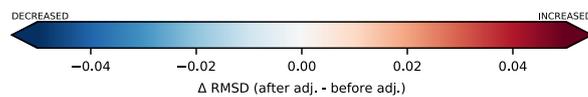
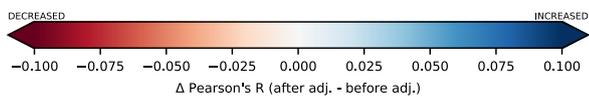


Fig. 38: Differences in Pearson's R (left) and RMSD (right) between the unadjusted and the adjusted versions of ESA CCI SM v04.5 COMBINED within SSP_6 (1987-07-09 to 1991-08-05) with ERA5 SWVL1 as the reference.

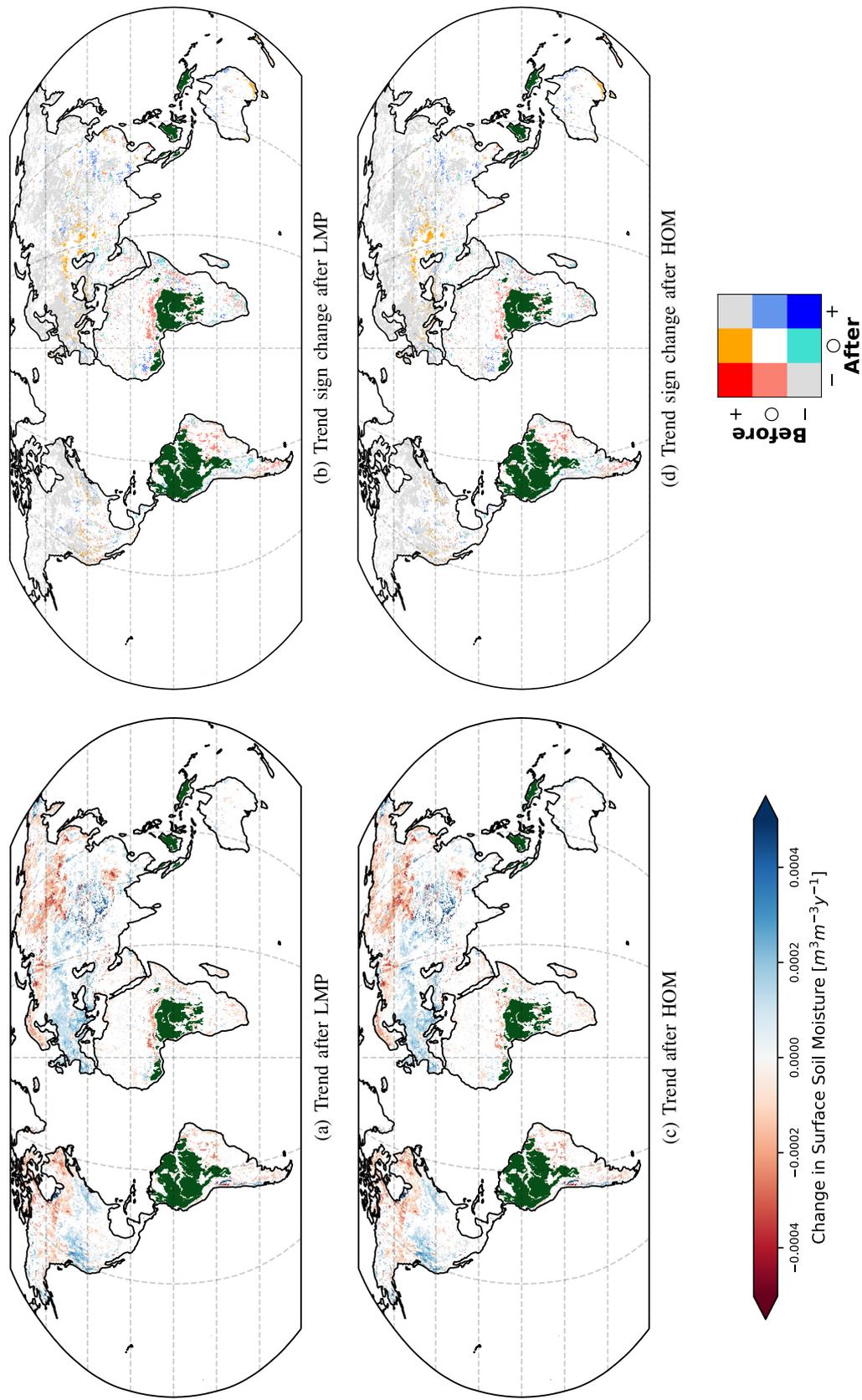


Fig. 39: Left: Significant ($p < 0.05$) trends (period 1991-2018) in ESA CCI SM after adjustment. Right: Sign change in significant SM trends after adjustment. "+" indicates wetting, "-" drying and "○" non-significant trends before, respectively after adjustment. Masked areas correspond to the ESA CCI SM dense vegetation mask (dark green).

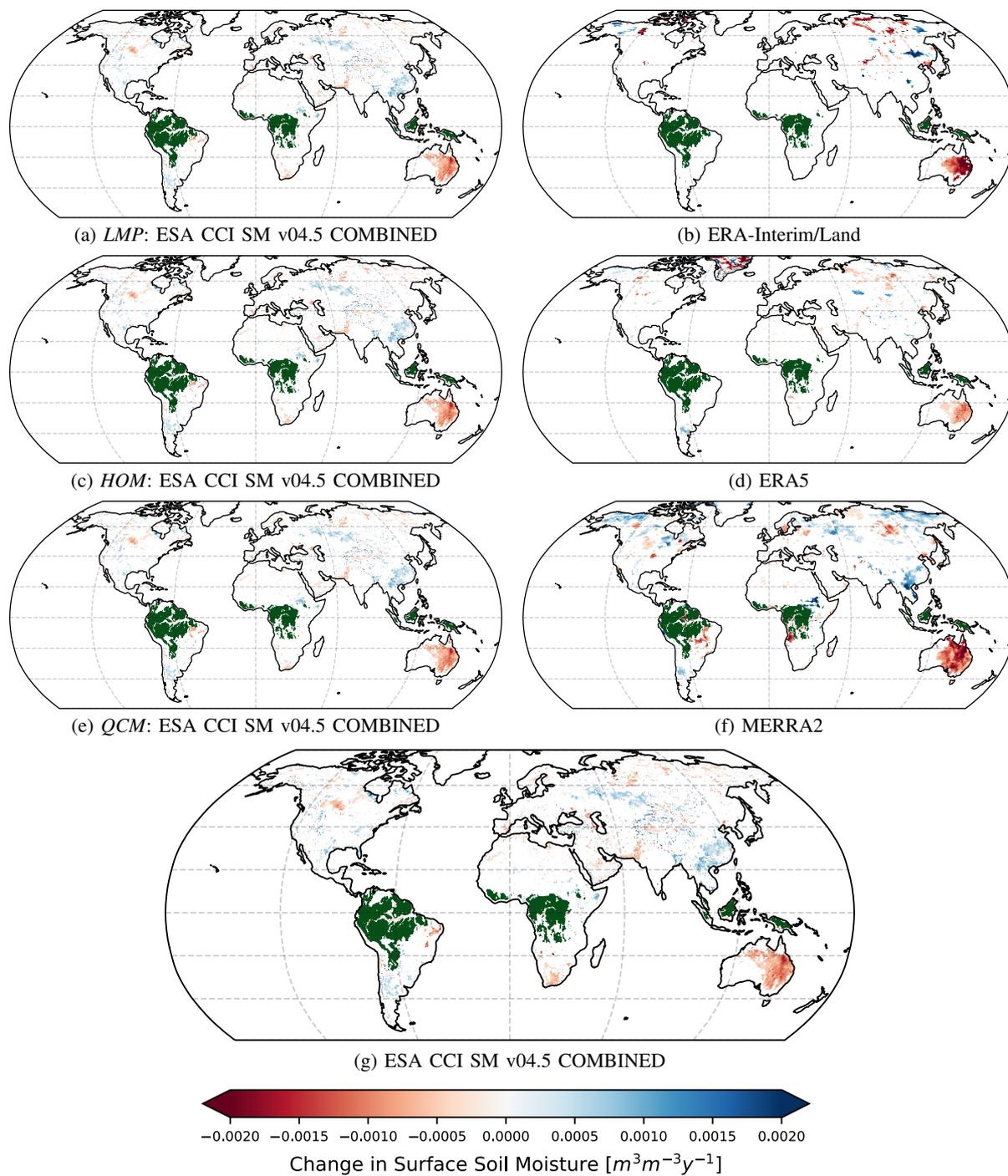


Fig. 40: Significant ($p < 0.05$) trends in the period 2010-01-05 to 2018-06-30 in the original (g), LMP adjusted (a), HOM adjusted (c) and QCM adjusted (e) ESA CCI SM v04.5 COMBINED data set and in ERA-Interim/Land SWVL1 (b), ERA5 SWVL1 (d) and MERRA2 SFMC (f); with areas masked out according to the ESA CCI SM dense vegetation mask.

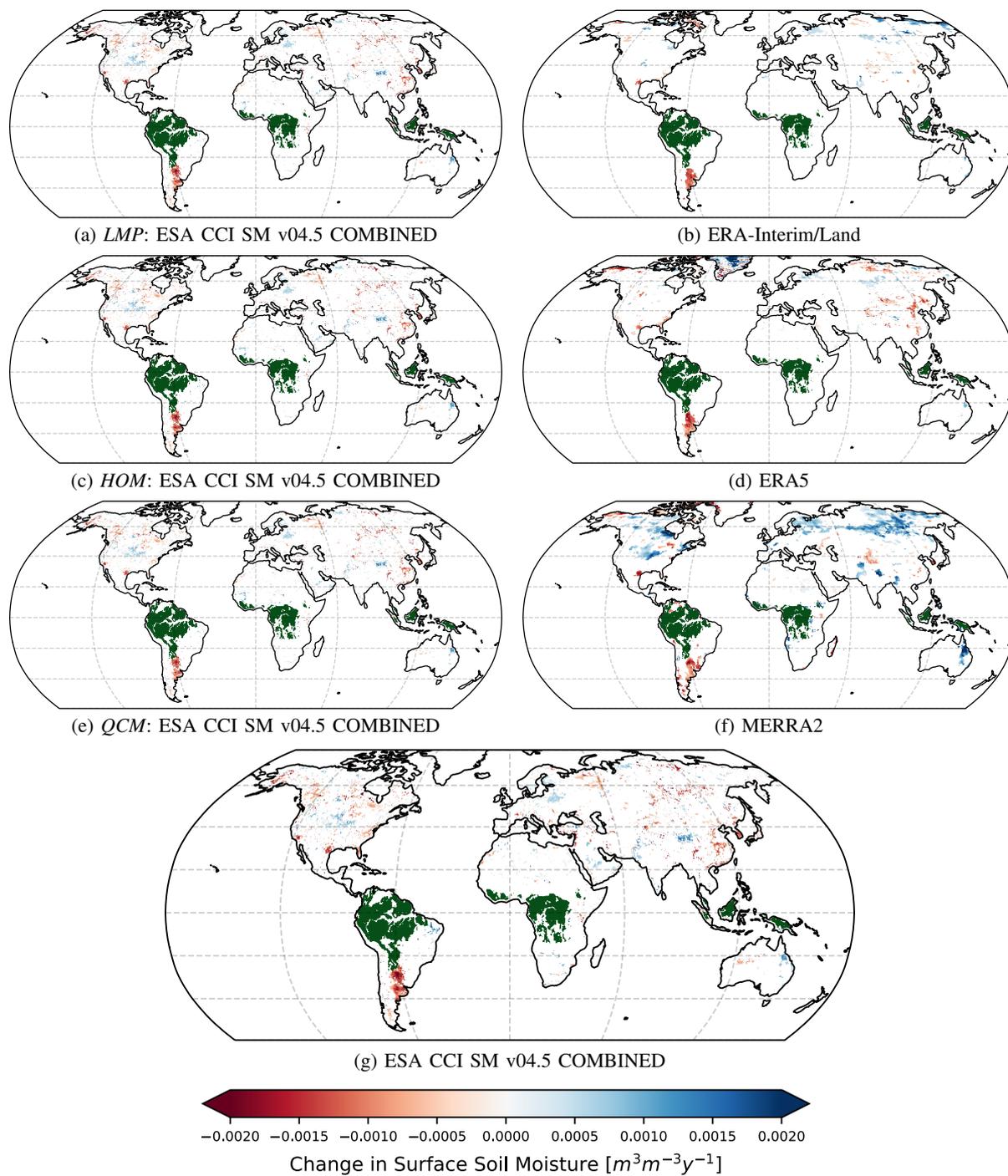


Fig. 41: Significant ($p < 0.05$) trends in the period 2002-06-19 to 2010-01-15 in the original (g), LMP adjusted (a), HOM adjusted (c) and QCM adjusted (e) ESA CCI SM v04.5 COMBINED data set and in ERA-Interim/Land SWVL1 (b), ERA5 SWVL1 (d) and MERRA2 SFMC (f); with areas masked out according to the ESA CCI SM dense vegetation mask.

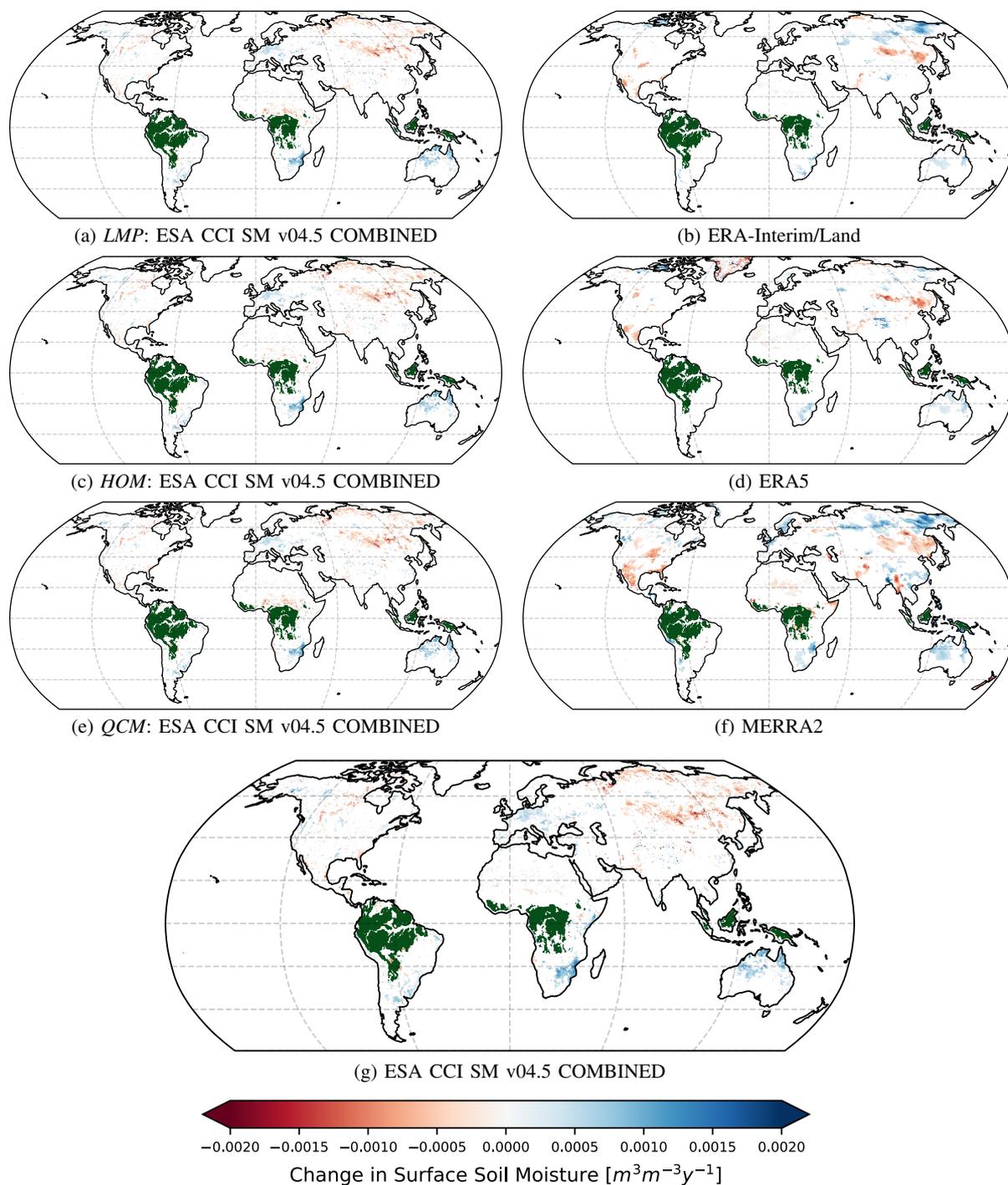


Fig. 42: Significant ($p < 0.05$) trends in the period 1991-08-05 to 2002-06-19 in the original (g), LMP adjusted (a), HOM adjusted (c) and QCM adjusted (e) ESA CCI SM v04.5 COMBINED data set and in ERA-Interim/Land SWVL1 (b), ERA5 SWVL1 (d) and MERRA2 SFMC (f); with areas masked out according to the ESA CCI SM dense vegetation mask.

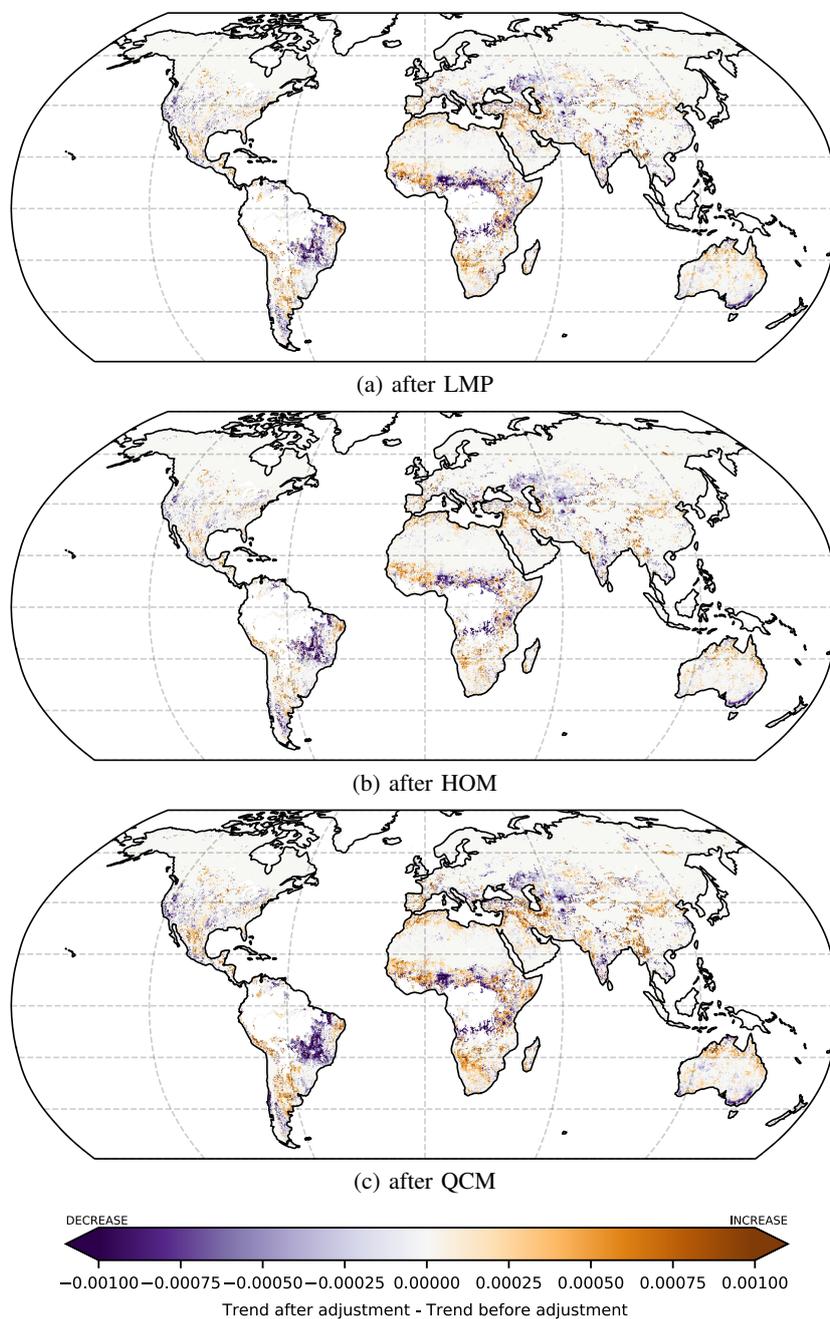


Fig. 43: Difference in significant trends in ESA CCI SM v4.5 COMBINED over the period 1991-2018 before and after correction with LMP, HOM and QCM.

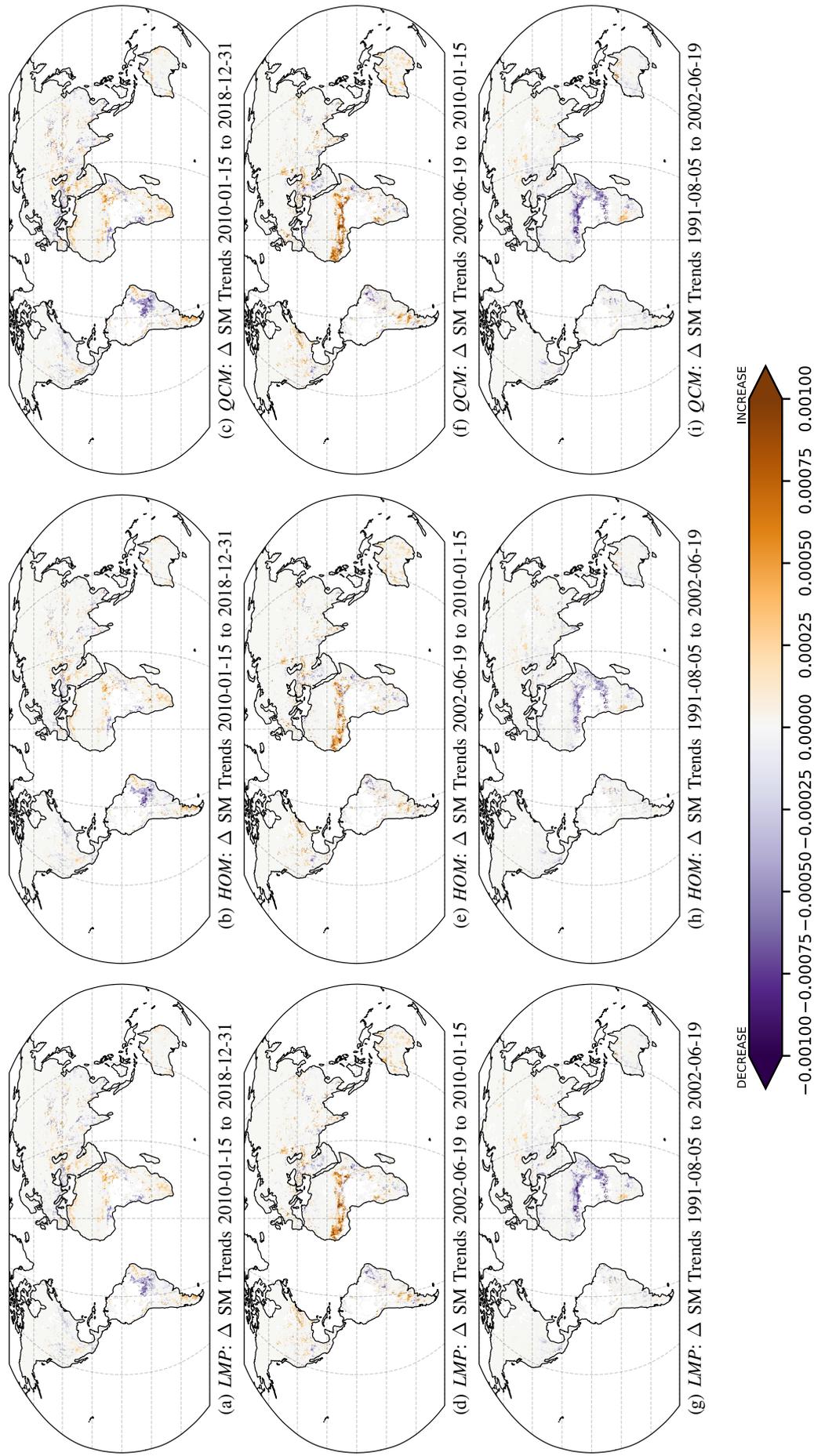


Fig. 44: Difference in significant trends in ESA CCI SM v4.5 COMBINED across three SSP pairs before and after correction with LMP, HOM and QCM.